Synthetic biology approaches for protein production optimization in bacterial cell factories

Rennig, Maja

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Synthetic biology approaches for protein production optimization in bacterial cell factories

PhD Thesis

Maja Simone Rennig

Novo Nordisk Foundation Center for Biosustainability
Technical University of Denmark

December 2017

Synthetic biology approaches for protein production optimization in bacterial cell factories

PhD thesis written by Maja Rennig

Supervised by Senior Scientist Morten H. H. Nørholm

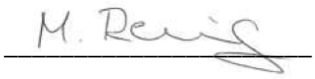
To my grandfather,

the person who inspired me the most in my life

and who I admire for what he achieved in his life.

Preface

This thesis is written as a partial fulfilment of the requirements to obtain a PhD degree at the Technical University of Denmark. The work presented in this thesis was carried out between October 2014 and December 2017 at the Novo Nordisk Foundation Center for Biosustainability at the Technical University of Denmark, Kgs. Lyngby, included a short research stay at Stockholm University in Sweden, and was supervised by Senior Scientist Morten H.H. Nørholm. Funding was provided by the Novo Nordisk Foundation.



Maja Rennig

Kgs. Lyngby, December 2017

Abstract

Society's strong dependence on fossil fuels and petroleum-based products leads not only to a rapid decline of natural oil reserves but contributes massively to global warming and environmental damage. This consequently urges society to look into more sustainable alternatives. Microorganisms present such sustainable alternative if converted into so-called microbial cell factories. Instead of crude oil, cell factories use renewable resources or waste products as source material. The challenge is, however, that microbial production needs to be economically feasible to compete with the classical chemical production. The development of a microbial cell factory typically takes up to 8 years of research and costs over $50 million. The production and selection of heterologous pathway proteins are major bottlenecks encountered in the construction of a cell factory. Thus, new approaches for the optimization of recombinant protein production and screening techniques with high capacity for the identification of the best performing enzymes are continually developed.

This thesis aims to equip researchers with a fundamental knowledge about protein biosynthesis necessary for the understanding of protein production bottlenecks. Moreover, the thesis guides through the possible causes of low protein yields and presents available approaches for optimization of the protein and the host.

The main work presented in this thesis provides and applies a new synthetic biology approach for the optimization and selection of recombinant proteins. A major bottleneck during production is translation initiation. By creating sequence libraries of the translation initiation region, protein production can be improved substantially in Gram-negative and Gram-positive bacteria. The design of versatile and tuneable translational coupling devices and their fusion to antibiotic selection markers enables subsequent selection of high-expressing constructs. The approach is a simple and inexpensive alternative to advanced screening techniques. In addition, a second synthetic biology approach provides the means for fast and efficient plasmid backbone swapping and is a versatile tool for the design and construction of optimal protein production constructs.

Dansk resumé

Samfundets stærke afhængighed af fossile brændstoffer og oliebaserede produkter fører ikke kun til et hurtigt fald i den naturlige oliereserver, men bidrager også kraftigt til global opvarmning og forurening. Dette motiverer derfor forskere til at undersøge mere bæredygtige alternativer. Mikroorganismer repræsenterer et sådant bæredygtigt alternativ, når de omdannes til såkaldte mikrobielle cellefabrikker. I stedet for råolie bruger cellefabrikker vedvarende ressourcer eller affaldsprodukter som udgangsmateriale. Udfordringen er, at mikrobiel produktion skal være økonomisk bæredygtig for at konkurrere med den klassiske kemiske produktion. Udviklingen af en mikrobiel cellefabrik tager typisk op til 8 år og koster $50 millioner. Produktionen af heterologe pathway proteiner samt den efterfølgende udvælgelse af gode kandidater er en flaskehals, der opstår ved konstruktionen af en cellefabrik. Nye tilgange til optimering af rekombinante proteinproduktions- og screeningsteknikker med høj kapacitet til identifikation af de mest effektive enzymer udvikles løbende.

Denne afhandling har til hensigt at give forskeren en grundlæggende viden om proteinbiosyntese, der er nødvendig for at forstå flaskehalse i proteinproduktionen. Desuden guider afhandlingen gennem de mulige årsager til lave proteinudbytter og præsenterer tilgange til optimering af proteinet og værtorganismen.

Hovedarbejdet i denne afhandling giver og anvender en ny syntetisk biologi tilgang til optimering og udvælgelse af rekombinante proteiner. En stor flaskehals under produktionen er translationsinitiering. Oprettelsen af sekvensbiblioteker kan forbedre proteinproduktionen væsentligt i Gram-negative og Gram-positive bakterier. Udformningen af alsidige og tuneable translationelle koblingskassette og deres fusion til et antibiotikaresistensgen muliggør efterfølgende udvælgelse af konstrukter med højeste ekspression. Tilgangen er et simpelt og billigt alternativ til avancerede screeningsteknikker.

Acknowledgement

Doing a PhD was definitively not a straight beautiful avenue, it was more like a bumpy hiking path going up and down with fixed rope routes in between, but in the end you reach the top and the view is just beautiful. On this hike I have met many amazing people that I want to express my gratitude to.

First of all, I want to thank my supervisor Morten Nørholm for believing in me and giving me the opportunity to pursue a PhD in his group. For the three years that followed I have to thank you for being an extremely enthusiastic supervisor full of ideas whom it was a pleasure to work with and not for! I have to thank you for always having an open door and time for questions, for appreciating my opinion and for never saying no to my own ideas. You always stayed positive and cheerful no matter if experiments were not working out or if we received an unreasonable review on a manuscript.

I also want to thank all current and former members of the Microbial Evolution and Synthetic Biology group: Karina Søgaard, Ulla Christensen, Victor Forman, Emil Fisher, Darío Vázquez Albacete, thank you for all the coffees, brainstorming and a wonderful group environment back then in the forest. Sofie Wendel and Mafalda Cavaleiro, thank you for so many fun moments in the lab, your positive attitude, advice and help and all the nice memories that we created outside of the lab. Thanks to the two amazing postdocs Virginia Martínez and Se Hyeuk Kim for being so patient and helpful. I also want to thank Tonja Hobel, Ida Lauritsen, Silvia Capucci, Alexey Dudnik and Roberto Ferro for an amazing group environment and all the laughs since we moved into the new building. Thanks for your help and advice, for listening to all my complaints and for cheering me up once in a while. A special thanks goes to Roberto Ferro for a very successful collaboration and to Cristina Hernández Rollán for being a great help in this project. I also want to thank Camilla Johansen, the best master student one can ask for.

At this point I would also like to thank my previous mentors Christine Oswald and Robert Ernst. Without your guidance, your strong belief in my skills and your positive attitude towards science, I would have never made it to the point where I am now. Robert, your once in a while check-up texts and calls meant a lot to me and I am grateful to know that I am always welcome in your lab. I also want to thank my collaboration partners in Stockholm. A special thanks to Daniel Daley and

Kiavash Mirzadeh for fruitful discussions on our projects. And thanks to the whole Daley lab for making my short research stay in Stockholm such a nice experience.

Working at CfB was a great experience. In the past three years I met so many smart, talented and inspiring people, all of whom I want to express my gratitude to. I want to thank Hemanshu Mundhada for our nice and still ongoing collaboration on the serine production project. At this point, I would also like to thank the CfB administration and all members of the JCC and PhD School Committee for appreciating my opinion and respecting my role as PhD Club chairman. A special thanks goes to the very active members of the PhD Club, Anne Sofie Lærke Hansen, Kira Sarup Lytzen, Alicia Lis, Kristian Jensen, João Cardoso and Ida Lauritsen for your enthusiasm, your ideas and your help.

CfB would not have been the same without all the friends I made during my PhD. First a special thanks to all the Bactory students for "adopting" me and making me feel home in Copenhagen from the very first day on. Thanks to Rosa Aragão Börner, Alicia Jiménez, Gheorghe Borja, Nabin Aryal, Mikkel Lindegaard, Klara Bojanovic Machado and many more for making it so easy to go to CfB every day. A very special thanks goes to Patricia Calero and Isotta D'Arrigo for all the support and listening, for all the good moments and fun we had together. No matter if skiing in Sweden or partying in Budapest, it was always a blast! I don't want to forget about my other Copenhagen friends, Steven Ocampo, Carsten Beese and Ludovica Hengeller.

A very special thanks goes to Christian Lieven for becoming such a big part of my life during my time here in Copenhagen. I am so thankful for having you around me everyday and for everything you do for me. For your patience, for your support, for encouraging me, for listening, for making me laugh and cheering me up, for dragging me away from work when it's enough, for amazing vacations and adventures... This PhD would have been so much harder without you!

I want to thank my friends back in Germany for not forgetting about me, always keeping in touch and for visiting once in a while. Thanks to Alina, Susi, Anna, Kristina and Jui, Michele, Katha, Dominik and all the others. A special thanks goes to Lotte, Anke, Jule, Bici and Lulu for keeping our festival tradition alive over all these years. I'm already looking forward to July!
Finally, I want to thank my family for their love, trust and support in any situation. Thanks to my parents and brother for encouraging me to move to Copenhagen and thanks for your visits, which

brought a little bit of home to Copenhagen. With you behind me, I feel like I can master every situation. I wouldn't be the person I am without you. Thanks to my grandparents for always supporting me, for being role models to me and for once driving the long way to Copenhagen to visit me. Dankeschön!

List of publications

1 Rennig M., Martinez V., Mirzadeh K., Dunas F., Röjsäter B., Daley D.O., Nørholm M.H.H. (2017) **TARSyn: Tuneable antibiotic resistance devices enabling bacterial synthetic evolution and protein production.** ACS Synthetic Biology, DOI: 10.1021/acssynbio.7b00200

2 Rennig M., Daley D.O., Nørholm M.H.H. (2018) **Selection of highly expressed gene variants in *Escherichia coli* using translationally-coupled antibiotic selection markers.** In: Jensen M., Keasling J. (eds) Synthetic Metabolic Pathways. Methods in Molecular Biology 1671, pp 259-268. Humana Press, New York, NY

3 Ferro R.*, Rennig M.*, Hernandez Rollan C., Daley D.O., Nørholm M.H.H. (2018) **A synthetic biology approach for selection of highly expressed gene variants in Gram-positive bacteria.** (Submitted to Microbial Cell Factories)

**These authors contributed equally*

4 Kim S.H., Cavaleiro A.M., Rennig M., Nørholm M.H.H. (2016) **SEVA linkers: a versatile and automatable DNA backbone exchange standard for synthetic biology.** ACS Synthetic Biology 5 (10), pp 1177–1181

Publications not included in this thesis:

5 Rennig M.*, Mundhada H.*, Nielsen A.T., Nørholm M.H.H. (2018) **An expression library based approach to improve L-serine production in *Escherichia coli*.** (Manuscript in preparation)

**These authors contributed equally*

6 Mirzadeh K., Elfageia R., Shilling, P., Cui H.L., Rennig M., Nørholm M.H.H., Daley D.O. (2018) **Synthetically evolved translation initiation regions enhance periplasmic protein production.** (Manuscript in preparation)

7 D'Arrigo I., Cardoso J., Rennig M., Sonnenschein N., Herrgård M.J., Long K.S. (2018) **Investigation of the *Pseudomonas putida* KT2440 transcriptome in different sole carbon sources reveals novel nutrient uptake systems.** (Manuscript in preparation for resubmission)

Patent applications not included in this thesis:

8 Rennig M., Ferro R., Mirzadeh K., Daley D.O., Martinez V., Lauritsen I., Nørholm M.H.H. **Translational coupling devices.** EP17186000.0 (filed August 2017).

Abbreviations and nomenclature

AbR	antibiotic resistance marker
acetyl-CoA	acetyl-Coenzyme A
ALE	adaptive laboratory evolution
aTc	anhydrotetracycline
ATP	adenosine triphosphate
B.C.	Before Christ
BAM	β-barrel assembly machinery
BCD	bicistronic design
CAI	codon adaptation index
CHO cells	Chinese hamster ovary cells
CO_2	carbon dioxide
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRP	cyclic AMP receptor protein
DNA	deoxyribonucleic acid
EF	elongation factor
FACS	fluorescence activated cell sorting
FDA	Food and Drug Administration
GDP	guanosine diphosphate
GFP	green fluorescent protein
goi	gene of interest
GPCR	G-protein coupled receptor
GRAS	generally regarded as safe
GST	glutathione S-transferase
GTP	guanosine triphosphate
His-tag	Poly histidine affinity tag
IF2, IF2, IF3	Initiation factor 1-3
MAGE	multiplex automated genome engineering
MBP	maltose binding protein
mRNA	messenger RNA
Mt	million tonnes
NGS	Next-generation sequencing
NusA	N-utilizing substance A
ori	origin of replication
PCR	polymerase chain reaction
POTRA domains	polypeptide transport associated domains
PPIase	cis/trans peptidyl-prolyl isomerase
RAGE	Cre recombinase-assisted genome engineering
RBS	ribosome binding site
RF1, RF2, RF3	release factor 1-3
RFF	ribosome recycling factor
RNA	ribonucleic acid
RNAP	RNA polymerase

rRNA	ribosomal RNA
SD sequence	Shine-Dalgarno sequence
SRP	signal recognition particle
SUMO	small ubiquitin-like modifier
synbio	synthetic biology
TALENs	transcription activator-like effector nucleases
Tat system	Twin-arginine translocation system
TF	trigger factor
TIR	translation initiation region
tRNA	transfer RNA
TrxA	thioredoxin A
TSS	translational start site
UN	United Nations
UTR	untranslated region
ZFNs	Zinc finger nucleases

List of Figures

List of Tables

Table of contents

Introduction and thesis outline

In today's world, the society strongly depends on petroleum-based products. Those are derived from crude oil and comprise gasoline, jet fuels and diesel fuels but also daily products, such as textiles, medicines and cosmetics[1]. In the Key World Energy Statistics from 2017 the International Energy Agency highlights trends in energy production and use[2]. They find that crude oil production has grown from 2,869 Mt in 1973 to 4,321 Mt in 2016. These numbers are indicative of a higher dependency on oil which is accompanied by a decline of natural oil reserves. However, dwindling oil supplies are not the sole problem society is facing: a petroleum-based industry also produces huge amounts of greenhouse gases contributing to global warming and environmental damage. In the same report CO_2 emission from fuel combustion increased from 15,458 Mt of CO_2 in 1973 to 32,294 Mt of CO_2 in 2016.

These negative consequences of petroleum-based sources are extremely worrying for our planet's future and urge society to look into more sustainable alternatives. This was highlighted at the 2015's General Assembly of the United Nations (UN) where the UN members adopted a set of sustainable development goals that should be achieved in the next 15 years[3]. Sustainable development means meeting the needs of today's society without compromising those of future generations[4]. Among other items, the assembly concluded that we need to ensure "sustainable consumption and production patterns" by "encourage[ing] companies […] to adopt sustainable practices"[3]. Here, the focus ought to be bio-based and sustainable production. Sustainable production describes a production process that minimizes waste output and pollution, by e.g. recycling, that conserves natural resources and minimizes energy input while still being economically feasible[5].

Among other sectors, pharmaceutical and chemical industries are looking for sustainable alternatives to produce medicine and chemicals, respectively[6]. Microorganisms present such sustainable alternatives that show great promise to replace petroleum-based production. Thus, dedicated research groups have made it their task to convert microorganisms into so-called cell factories for production of high-value compounds, such as medication, but also bulk chemicals, biofuels and -plastics, using renewable resources or waste products. For thousands of years, microorganisms have already been used in the production of beer, bread and cheese. More recent

examples of microbial production include the production of recombinant human insulin in *Escherichia coli* and *Saccharomyces cerevisiae*[7], the production of opioids and the anti-malarial drug artemisinin in *S. cerevisiae*[8,9], the production of therapeutical proteins and biofuels in *E. coli*[10,11], the production of industrial enzymes in *Bacillus subtilis*[12] or the production of industrial compounds or therapeutical proteins in *Lactococcus lactis*[13].

However, the challenge here is that microbial production needs to be economically feasible to compete with classical chemical production[14]. This requires a long process of developing and optimizing a microbial cell factory. Optimization and development can be seen as a cycle going through analysis and engineering of the strain, which is called metabolic engineering[15]. Very often protein production represents the major bottleneck in this process.

The work conducted in this PhD thesis aims at optimizing protein production in bacterial cell factories by using synthetic biology approaches. For this reason, the thesis focuses on tool development for protein production optimization that can be applied in the engineering part of the metabolic engineering cycle. The thesis is divided into three introductory chapters, where concepts of synthetic biology, cell factory design and engineering and protein production, which are important for the understanding of this thesis, and protein production optimization in general, are explained.

Chapter 1 gives an introduction to the field of synthetic biology and its impact on recombinant protein production. Furthermore, the chapter introduces the concept of a cell factory and the metabolic engineering Design-Build-Test paradigm, and shows how protein production optimization can be part of it.

Chapter 2 comprises the fundamental biological processes of protein biosynthesis including transcription, translation and finally protein folding and targeting. A comprehensive knowledge of these fundamental processes is necessary to understand engineering approaches for protein production optimization.

In Chapter 3 the importance of recombinant protein production in the creation of cell factories is elucidated, its bottlenecks and several techniques and approaches for optimization are introduced.

The need for fast and simple universal technologies for the optimization of protein production in bacteria fuelled the work presented in this thesis. Papers 1 to 4 detail the work conducted in this regard in form of submitted manuscripts or published articles.

In this thesis, an experimental tool for the optimization of protein production was developed. Utilizing translation initiation sequence libraries in combination with an antibiotic selection-based screening technology, protein production levels have been raised substantially (Paper 1 and 2). This method was initially implemented for the Gram-negative bacterium *E. coli* but in addition was successfully applied to the Gram-positive hosts *B. subtilis* and *L. lactis* (Paper 3). Paper 4 contains a published article introducing a synthetic biology standard for the rapid exchange of parts. This tool demonstrates to be a useful for protein production optimization. It allows for an easy exchange of elements that affect protein production, such as gene copy number or the burden of an antibiotic resistance marker. Two appendices give short insights into the application of the first synbio approach for the optimization of expression of an operon and for the optimization of expression on the genome of *E. coli*.

Chapter 1: A new era of biotechnology

Biotechnology is an interdisciplinary field of science that describes the application of engineering principles to biological organisms and processes. Biotechnology dates back several thousands of years when the yeast *Saccharomyces cerevisiae* was used in fermentation processes for the production of beer, wine and bread, as well as when people produced cheese and yoghurt with the help of lactic acid producing bacteria. In the beginning of the 20th century Charles Weizmann established the field of industrial biotechnology by producing acetone and butanol at large-scale by fermentation of the bacterium *Clostridium acetobutylicum*[16]. About 20 years later, in the 1940s, some years after its discovery, the mass production of penicillin in microbes demonstrated the potential of biotechnology in the medical field. Later on, the production of synthetic human insulin in *Escherichia coli* paved the way for its medical and pharmaceutical application[17]. In the 1980s synthetic human insulin became the first approved genetically engineered pharmaceutical product[18,19].

Since the beginning of the 21st century the field of biotechnology expanded quickly. 20 years after the launch of synthetic human insulin by one of the first biotech companies, Genentech[20], more than 1500 biotech companies existed in Europe and about 1300 in the US[21]. The negative consequences of the petroleum-based industry, such as global warming and environmental damage, urge scientists to focus on more sustainable alternatives. Therefore, modern biotechnology branches into many different areas, not only in the before mentioned industrial and medical field, but also in the environmental, agricultural and marine area. All fields have a common interest in harnessing biological processes to improve living conditions on our planet for today's and future generations. This includes reduction of our environmental footprint, the use of clean fossil-independent energy, e.g. biofuels, and the mitigation of the green house effect[22,23]. It also comprehends the cheap production and fast development of efficient medication for healthcare improvement, tackling also problems such as antibiotic resistance. It includes ensuring sufficient food supply by increasing yield and enhancing nutrition content. The general use of renewable resources and waste products for the production of such is anticipated[6].

1.1 A brief history of synthetic biology and its impact on recombinant protein production

Synthetic biology plays an important role in this new era of biotechnology and emerged in parallel with the latter. Consensus on a precise definition of synthetic biology has not yet been reached but it is generally agreed that it originates from the idea that engineering approaches can be harnessed to study and manipulate cellular systems. The study and manipulation of cellular systems has also substantially influenced the field of recombinant protein production (Figure 1A)[24]. And *vice versa*, detailed knowledge about the central processes of protein biosynthesis enabled successful manipulation of cellular systems.

The discovery of transcriptional regulation in the *lac* operon, that was described in detail by Jacob and Monod in 1961[25], pointed towards a possibility of programmed gene expression. With inventions such as molecular cloning[26–28] and DNA amplification by polymerase chain reaction (PCR)[29] this vision got closer to reality. In 1977, the first recombinant human protein, somatostatin, was synthesized in *E. coli*[30]. When in the 1990s, the process of DNA sequencing became automated[31] and together with the development of computational tools enabled whole genome sequencing[32,33], the potential to manipulate microorganisms and to generate recombinant DNA was obvious[34].

Advances in systems biology enabled the examination and engineering of organisms in computational networks[35,36]. Data for those networks was collected with high throughput techniques that analyzed protein, lipid, metabolite and RNA content of cells. These studies are generally entitled as 'omics' studies. 'omics' is a neologism that comprises scientific studies ending in –omics, such as genomics, proteomics, lipidomics, metabolomics or transcriptomics. Following a computational analysis, cellular processes can be rationally manipulated[34,37] (Figure 1B). The advent of the omics era constituted an equally important driving force for recombinant protein production optimization. The sudden availability of thousands of protein sequences with unknown function demanded new approaches to produce those proteins in high quantity and quality for functional and structural characterization[38].

In parallel, standard molecular parts for regulating biological networks were developed[34]. In the beginning of the 21th century, the first synthetic genetic circuits were constructed[34,39,40]. Those circuits were inspired by electronic circuits and were designed to perform logical functions. The

workflow to build these circuits became a characteristic feature of synthetic biology tool development[34]: after an initial design phase, the circuit is physically constructed, and following experimental measurements, the system is adjusted to fit the hypothesized characteristics. This process is similar to the Design-Build-Test engineering paradigm described in section 1.2.1 and was applied for the tool development in this PhD thesis.

To simplify the design and evaluate the performance of new circuits a standardized way of storing and assembling parts was anticipated[34]. The Registry of Standard Biological Parts (RSBP) was developed to catalogue those genetic parts as 'BioBricks'[34,41,42]. However, the development of new one-step DNA assembly methodologies, such as Golden Gate cloning[43], Gibson assembly[44] or Uracil Excision Cloning[45,46], made standardization of parts assembly difficult, which is the reason why part registries such as RSBP now serve more as a sequence databases[34]. The development of new DNA assembly methods also influenced the field of recombinant proteins production allowing for more rapid construction and higher throughput.

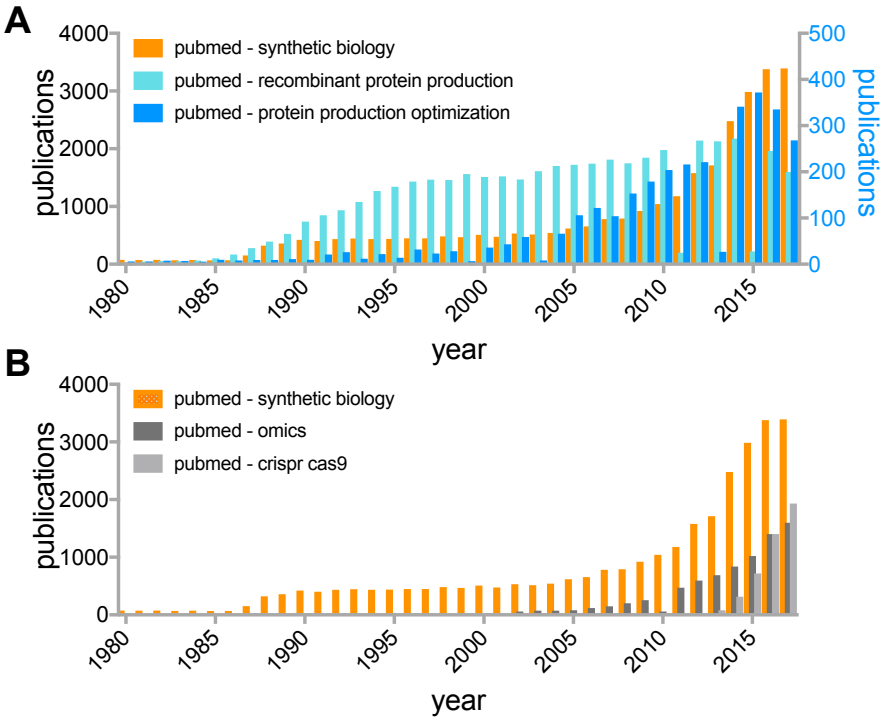


Figure 1: Comparison of the developments in synthetic biology and recombinant protein production judged by publication numbers. (A) Developments in synthetic biology are compared with recombinant protein production in general and with protein production optimization in specific. (B) The developments in the field of omics and the implementation of CRIPSR/Cas9 are compared with the development of synthetic biology.

However, synthetic biology anticipated to go beyond circuit engineering and envisioned characterization and engineering of whole organisms, to convert them into cellular devices capable of diagnosing diseases, fixing genetic defects or cleaning up environmental pollutants[34,47]. The rise of genome engineering technologies[48] as well as advanced DNA-assembly techniques[43–45] and a dramatic price-drop in DNA synthesis and sequencing[49], paved the way towards those future goals[34]. The emergence of *de novo* DNA synthesis also revolutionized recombinant protein production. Suddenly, restrictions to source material, such as pre-existing DNA templates, were removed and the possibility to synthetically design genes and to create novel functions and circuits arose[24].

In 2010, Gibson *et al*., synthesized and assembled a 582 kb genome for a viable *Mycoplasma* cell[50,51]. In another demonstration of genome-scale synthetic biology, Pósfai *et al*. reduced the genome of *E. coli* by 15 % resulting in a strain with enhanced electroporation efficiency and plasmid stability[52]. Likewise, the reduction of *L. lactis* genome had beneficial effects on recombinant protein production levels[53].

A platform for multiplex automated genome engineering (MAGE), which was developed in the Church lab[54], forms the basis for rapid and small modification, such as amino acid replacements, RBS modifications or promoter exchanges, in multiple loci of the *E. coli* genome. Homologous recombination provides the basis for larger genomic modifications on the genome, however, its efficiency is especially low in *E. coli*, the model bacterium for synthetic biology and recombinant protein production. Thus, systems from other organisms, such as λ-red recombination system, are supplied to facilitate engineering efforts in the *E. coli* genome[55,56] (see chapter 3.2). Nevertheless, the low efficiency requires the simultaneous insertion of a selectable marker, which cannot be removed scarlessly afterwards. Alternatively, Cre recombinase-assisted genome engineering (RAGE)[57] or *I-SceI* endonuclease cleavage can be used to enhance site-specific recombination[58]. Zinc finger nucleases (ZFNs)[59] and transcription activator-like effector nucleases (TALENs)[60] revolutionized genome editing in mammalian cells[61,62]. A few years later, the application of the CRISPR/Cas9 system for genetic manipulations started a "new era of genome integration"[63]. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and CRISPR-associated genes (Cas9) are essential in the bacterial adaptive immunity. The system can be hijacked to achieve site-specific DNA recognition and cleavage through the programmable endonuclease Cas9[64,65]. CRISPR/Cas9 assisted λ-red recombineering and MAGE enable the construction of desired strains

scarlessly and in a very short time frame[66–68]. Acceleration of genome engineering enabled easy manipulation of living cells to optimize protein folding, secretion and enable post-translational modifications (see Chapter 3.5). With those new technologies available the construction of a minimized genome can be done within months instead of years[69]. Despite all advances contributing to genome engineering, Pines *et al.* claim that with the decreasing costs of DNA synthesis, one may consider synthesizing a complete genome rather than having multiple engineering steps[70].

In recombinant protein production setups, transcriptional control through transcription factors and the use of different promoter sequences preceded the synbio era[71], although more recent approaches in synthetic biology extent the engineering possibilities to optimize protein production[24] (Figure 1A). Engineered transcription factors represent useful sensors and switches for metabolic engineering and protein production. They facilitate the design of new synthetic gene circuits that can e.g. increase population homogeneity in protein production setups. Recently, a recombinase feedback loop for robust protein expression was introduced supporting homogenous expression in a population[72]. Additionally, novel synthetic promoters or promoter libraries can be designed to control and tune gene expression[24]. This can also be done using RNA polymerases, so that toxic effects are prevented[73,74]. Toxicity often occurs when the cell cannot handle the protein load resulting in aggregation. The Stephanopoulous laboratory demonstrated that engineering of the endogenous *E. coli* RNA polymerase led to higher tolerance and increased protein production titers[73]. In another recent publication, artificial small RNAs regulate the T7 RNA polymerase and thus enhance soluble recombinant protein production in *E. coli*[75].

On the level of translation, creating synthetic RBS libraries allows researcher to tune translation rates. With this approach, optimal expression constructs could be identified in many cases[54,76]. Apart from that, synthetic biology approaches enable the engineering of the translated content. For instance, the engineering of aminoacyl tRNA synthetases allows for the incorporation of unnatural amino acids into a protein sequence[77,78]. Common applications of this synthetically expanded genetic code are light-induced crosslinking for protein-protein interactions or conformational changes, but also the light-activation of biochemical pathways through photocaged enzymes[79]. Moreover, unnatural amino acids can be incorporated in directed protein evolution experiments for novel or enhanced functions and, specifically immunogenic amino acids, in the synthesis of therapeutic proteins[80]. Liu *et al.* demonstrated the use of unnatural amino acids as transcriptional

on- and off-switches in *E. coli*[81], a useful tool for homogenous protein production and regulation of metabolic pathways. In general, however, engineering of the ribosome appears to be necessary to make incorporation of unnatural amino acids an efficient process. To not interfere with cell viability, Rackham and Chin created an orthogonal ribosome which only translates orthogonal messenger RNA and does not crosstalk with the native translation processes[82]. Ribosome engineering advances were recently reviewed by Liu *et al*.[83].

1.2 Metabolic engineering

Metabolic engineering is the science in which the engineering principles of synthetic biology, e.g. control circuits, are combined with decades of basic research to enable sustainable production of products extending from fuels and chemicals to food, feed and pharmaceuticals[34,84]. The natural metabolic network of a cell is generally not evolved for those practical applications. Therefore, cellular processes need to be modified and improved for certain needs[15]. Engineering of cellular processes can be either hypothesis-driven running through a Design-Build-Test cycle, which is elaborated on in 1.2.1, or evolution can be harnessed to reach an anticipated goal (section 1.2.2). Metabolic engineering of a protein production host, e.g. the regulation of carbon fluxes or the prevention of acetate accumulation, has proven to increase protein production levels. The over-expression of a recombinant protein interferes with the metabolism of a cell, e.g. by over-consuming the materials and energy coming from the TCA cycle, which finally results in an accumulation of acetate[85].

1.2.1 Design-Build-Test cycle

The construction of a cell factory is a time consuming and laborious process that iteratively cycles through designing, building and testing a certain metabolic strategy (Figure 2). Despite all advances over the years, the development of a new cell factory requires typically 6 to 8 years of research and over $50 million[84].

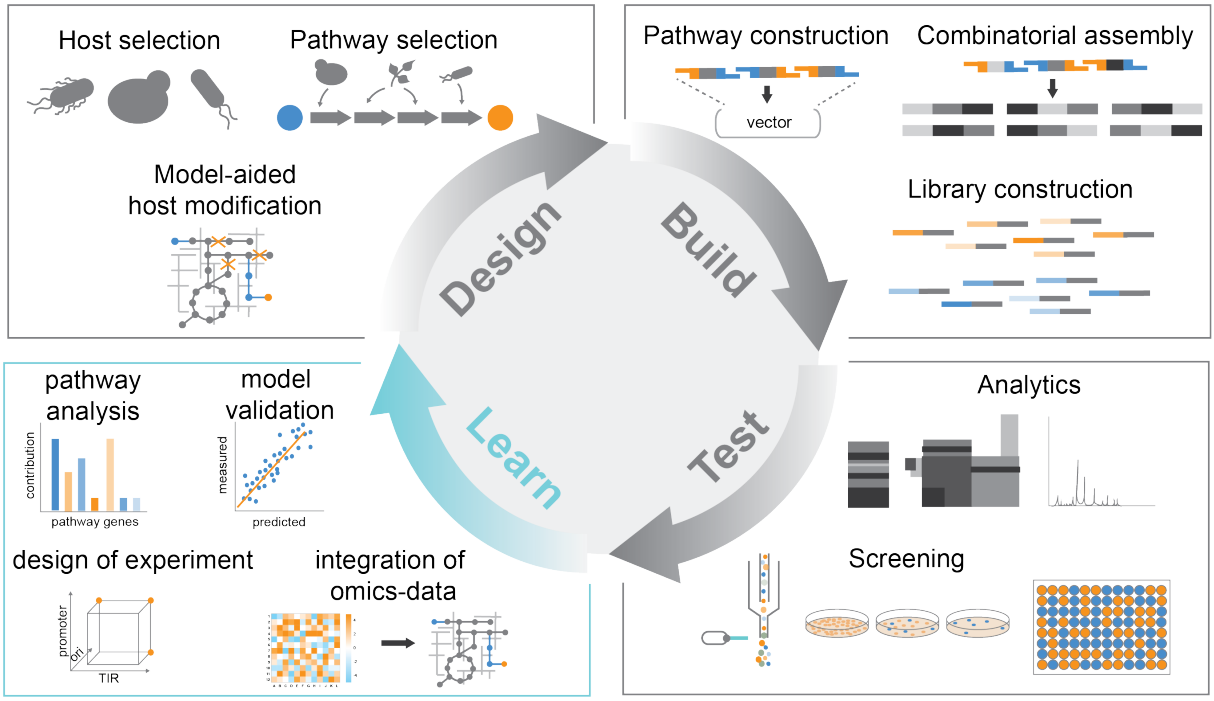


Figure 2: The Design-Build-Test-Learn cycle. In the design phase, the host, pathway and anticipated modifications are defined. In the build phase, the pathway is assembled. With screening approaches and analytical methods, the performance of the cell factory is evaluated in the test phase. In the later introduced learn phase, the data is collected and evaluated to improve the design process. (inspired by Petzold *et al.*[86])

The design phase

In the design phase, strategies are developed that can yield in a cell factory process. Here, appropriate strains and enzymes are identified. Hosts are usually selected based on required process conditions (temperature, pH, salt, solvent, etc.), strain physiology (growth-rate, product toxicity, metabolic versatility, etc.), feed stock and the availability of genetic engineering and gene expression tools at hand. The latter makes *Escherichia coli* and *Saccharomyces cerevisiae* favourite hosts, but recent genetic method developments have opened the option to use less studied hosts that possess desirable traits, such as product tolerance, or unique metabolic pathways[87]. Still, we are a long way from having synthetic biology tools that performs robustly in non-model organisms[88].

Similar requirements as for the host apply to the choice of enzymes. Advances in DNA synthesis no longer restrict researchers to using genetic material from natural sources, which provides great flexibility in the choice of enzymes[24]. These should perform well under the needed process conditions (in terms of temperature, pH, salt concentrations), enable rapid utilization of low-cost feedstocks and enable inexpensive product separation through e.g. secretion of the final product[88]. Even the apparent simpler case of enzyme production sometimes requires manipulation of many

genes. Identification of suitable proteins is mostly achieved by querying literature and databases (e.g. BRENDA, KEGG, EcoCyc or MetaCyc)[89,90]. Often, complete biosynthetic pathways are transferred from the original host to a heterologous host[87]. This has proven successful for the production of complex biochemicals such as opioids and artemisinic acid[9,91–93]. However, if there is not enough information on the biosynthetic pathway enzymes, retro-biosynthesis and pathway prediction algorithms, e.g. RetroPath or GEM-Path, can help to bridge the gaps and identify *de novo* production pathways[94,95]. Nevertheless, retro-biosynthesis is not further elaborated on here as this thesis focuses on protein production optimization. Retro-biosynthesis is reviewed in detail elsewhere[96]. To combine variables, e.g. promoter sequences or gene sequences, most efficiently, it should be standard practice to implement *in silico* tools, such as Design of Experiment (DoE) methods, in the design phase[97].

The use of stoichiometric models, such as genome-scale metabolic models can be very useful in the design process. Those models can be used to predict metabolic fluxes and how they are affected by genetic manipulation of the host[98]. Algorithms that predict the effects of gene over-expression, gene or reaction knockouts and insertion of heterologous pathways can be combined with genome-scale metabolic models[99]. So-called ME-models (for metabolism and expression) constitute the newest generation of stoichiometric models, accounting for protein biosynthesis, e.g. transcription and translation processes[100,101]. Thanks to the on-going development of computational tools, the design phase is very efficient and throughput is very high. After the required manipulations to the host are identified, the strategies and genetic elements to achieve them need to be designed. The choice of DNA assembly techniques and genome-engineering methods will determine the pace of the building phase.

The build phase

In the build phase, the previously identified manipulations, e.g. integration of heterologous enzymes, gene knockout or over-expression, are incorporated into the host. Advances in DNA synthesis and genome integration allow for the generation of large libraries[84,102,103].

Despite the successful integration of heterologous pathways, genes need to be expressed. Gene expression can very often be a major bottleneck in metabolic engineering but the development of tools for production optimization is progressing. With the ability to vary promoter[104–106] and translation initiation strength[107,108], to optimize stability of the mRNA[109,110] and to prevent the

resulting protein from aggregating, there are many starting points to alter protein production. Chapter 3 discusses in more details the bottlenecks and optimization approaches in protein production. Paper 4 of this thesis provides a plasmid backbone series, which allows for fast and easy parts exchange and provides the means to optimize production of proteins and biosynthetic pathways. We see one application of this toolbox, in the construction of a microbial cell factory.

To overcome physiological limitations, such as product toxicity, enzyme toxicity or insufficient substrate utilization, evolutionary approaches show much more capacity than rational engineering. Therefore, adaptive laboratory evolution (ALE) can be applied in the build phase. ALE is described in more detail in section 1.2.2.

The test phase

The performance of the constructed strain is evaluated in the test phase. Available screening approaches are described in more detail in section 3.6. High-throughput screening approaches are preferred to analytical approaches, as the latter are very time-consuming and expensive. Furthermore, the relatively low throughput of analytical methods is not applicable for the screening of very large libraries. In this thesis, a high-throughput screening tool is developed, which provides the means to be applied in the test phase of the construction of a microbial cell factory (Paper 1-3).

Apart from protein production and product formation, growth rate, tolerance and mutations will be evaluated with the help of omics tools: (i) next-generation sequencing (NGS) permits comprehensive validation of the genome of the engineered strain, (ii) RNA-seq analysis can be employed to identify up- or down-regulated genes, unintended mutations and other transcriptional failures, and (iii) proteomic assays provide means to quantify of protein levels over time, identify pathway bottlenecks and characterize synthetic parts[86].

The learn phase

The learn phase is not always considered a part of the Design-Build-Test cycle. However, it is perhaps the most important step to increase the rate of success. With the appearance of efficient computational and high-throughput molecular biology techniques, a step needed to be introduced to compute the large pool of generated data after each iteration, which can then be used to improve the design and build process in the next iteration[88,111]. It is a robust practise to avoid mistakes or

unnecessary steps. Generated omics data can be integrated into genome-scale metabolic models and statistical models allow for the analysis of complex data. This data can be used to identify trends or patterns that can be helpful for the re-design[84]. As more and more data is collected, the necessity of using sophisticated data-analysis tools, such as machine-learning algorithms for, e.g. pathway prediction or automated DoE, becomes more and more important. Being one of the first of its kind, a data-driven statistical prediction tool for membrane protein expression levels from nucleotide sequence was recently developed[112].

1.2.2 Adaptive laboratory evolution (ALE)

Once a metabolic pathway has been constructed, in most cases further optimization is still necessary to increase yield. For this reason, adaptive laboratory evolution (ALE) can be used[113]. Here, the organism adapts to the controlled conditions that it is exposed to. Over time stable mutations leading to a beneficial phenotype may accumulate[113]. NGS can then be used to identify which mutations are responsible for the optimal phenotype[114]. ALE can either be performed in serial dilutions or in continuous cultures[113]. When using serial dilutions, cells are propagated in batch cultures. Each time that the culture reaches a certain density, an aliquot is transferred to a fresh culture. This iterative process is repeated until cells reach the desired phenotype, e.g. a desired growth rate[113]. Alternatively, chemostat cultures have the advantage that conditions such as pH, oxygen levels and cell density are kept constant[113]. Organisms such as bacteria and yeast are normally subjected to ALE as their fast generation time allows for rapid evolution[113,115].

ALE has proven to be very efficient for the optimization of growth on non-preferred carbon sources, such as glycerol for *E. coli*[116] and xylose and galactose for *S. cerevisiae*[117,118]. ALE is also very useful in the adaption to environmental stresses such as elevated temperatures[119,120] or osmotic stress[113,121], or in the adaptation to toxic compounds. For example, in the production of biofuels, *E. coli* needs to obtain tolerance towards organic solvents such as ethanol[122,123] or isobutanol[124,125].

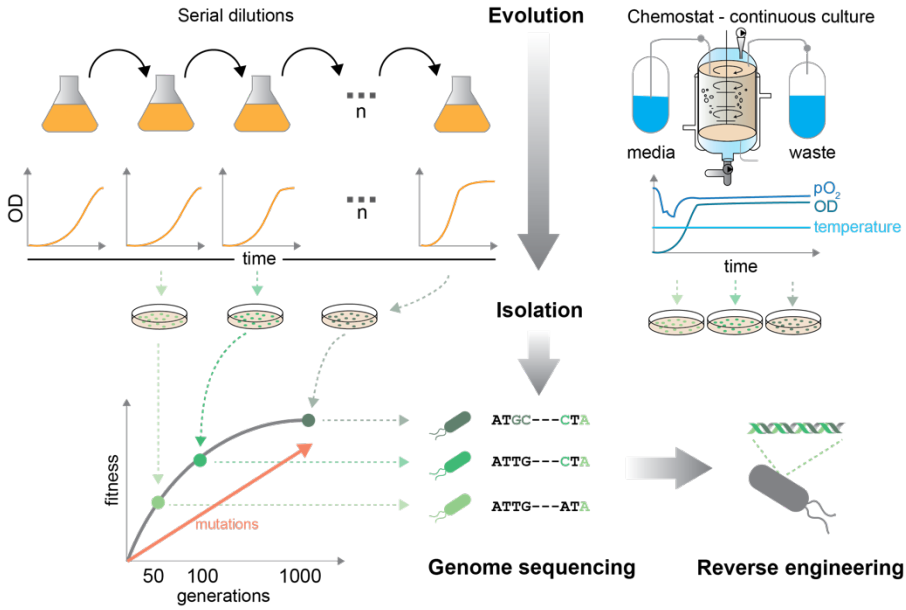


Figure 3: Adaptive Laboratory Evolution (ALE). Strain evolution can either be performed in a serial dilution experiment or in a continuous culture in a chemostat. Over time, fitness of the strains will increase as will the number of mutations occurring. Evolved clones are isolated and their genome sequenced. Identified mutations will be reverse engineered into the original strain to confirm their beneficial effect. (inspired by Dragosits and Mattanovich[113]).

In a recent engineering effort, the tolerance of *E. coli* towards L-serine was improved by ALE. L-serine is a very attractive product in fermentation as its theoretical yield from glucose is the highest among the amino acids but its high toxicity limits microbial production[126,127]. The evolved strains produced 37 g/L of L-serine, one of the highest reported titers, and improved productivity by 4-fold compared to the unevolved strain[126,127]. This tolerant strain was used to further optimize L-serine production in *E. coli* by applying an expression library-based optimization and selection approach which was developed and applied in Papers 1-3 (see Appendix 1). However, as Atsumi *et al*. previously showed in the case of isobutanol, developing tolerance towards a product does not necessarily mean an increase in production titers[125]. Tolerance could also be achieved by degradation of the product.

1.3 Relevant microbial cell factories

The choice of host is an important step in the development of microbial cell factories as well as for protein production setups. Two microbial cell factories dominate the market. *Escherichia coli* and *Saccharomyces cerevisiae* are by far the most studied organisms with a whole palette of engineering tools available and a collection of established products in pharmaceutical, food and biotechnological industry. Filamentous fungi, including *Aspergillus*, *Trichoderma*, *Penicillium* and

Rhizopus species, are preferred industrial hosts for recombinant protein production. Likewise, *Bacillus subtilis* is a profitable industrial host for protein production thanks to its ability to secrete large amounts of proteins. *Lactococcus lactis* is an increasingly attractive host for the production of plant-based proteins, secondary metabolites and membrane proteins. As *E. coli* and the latter two Gram-positive hosts have been used extensively in the studies included in this thesis, a brief description of the three hosts will follow. Other selected hosts are summarized in Table 1.

1.3.1 *Escherichia coli*

The Gram-negative, facultatively anaerobic, rod-shaped bacterium *E. coli* is probably the best studied and most widely used organism[128]. *E. coli* seems to be the ideal candidate for metabolic engineering as it has a fast doubling time, possesses metabolic versatility, both, a well-characterized genome and metabolism and plenty of genetic tools for strain manipulation[128]. Altogether, *E. coli* has a very long history of use and is a very cheap host for production. Even though some *E. coli* strains might be pathogenic, most are part of the natural human gut flora and its products are regarded as safe[10]. About 30% of recombinant biologics on the market are manufactured in *E. coli*[10,128]. FDA approved recombinant protein drugs, such as the interferon β-1b Extavia (Novartis, 2009) used in multiple sclerosis treatment[10], amino acids, bio fuels, organic solvents and polymers are produced in *E. coli*[11]. Some major drawbacks are the lack of post-translational modification machineries which carry out common modifications to human proteins[129], its poor secretion capacity[130] and the production of endotoxins[128]. Endotoxins trigger a strong immune response in mammals. They are as relatively thermostable lipopolysaccharides part of the cell envelope and are mostly released upon cell lysis.

1.3.2 *Bacillus subtilis*

Bacillus subtilis is a Gram-positive, facultatively anaerobic, rod-shaped bacterium found mostly in soil[12]. Some genus can form endospores and thus survive in extreme environmental conditions[131]. *B. subtilis* is non-pathogenic and does not produce any endotoxin leading to the GRAS status by the FDA, which means that *B. subtilis* can be used during food production[12]. Currently, *B. subtilis* is the best-characterized Gram-positive bacterium. It is a very attractive organism for protein production for several reasons. First of all, Gram-positives contain only one membrane, which simplifies protein secretion[132]. Secondly, due to an intermediate G/C-content, there is no pronounced codon bias. And lastly, towards the end of exponential phase *B. subtilis* becomes

naturally competent and generally performs well in homologous recombination, which facilitates genome engineering[133]. Similar to *E. coli*, *B. subtilis* has a fast doubling time, possesses metabolic versatility and a well-characterized genome and metabolism[134]. Despite the recent development of engineering tools like CRIPSR/Cas9 systems and expression platforms[106,135,136] there is still a distinct lack of genetic tools for strain manipulation.

1.3.3 *Lactococcus lactis*

Lactococcus lactis is a Gram-positive, facultatively anaerobic, spherical-shaped, non-sporulating bacterium that was originally isolated from plant[13]. However, it is mostly associated with dairy products[13]. *L. lactis* has been used for centuries in food fermentations, especially cheese, yoghurt, pickled vegetables and the like. Because of this, *L. lactis* has GRAS status[13]. Over the past decades, *L. lactis* has become the model lactic acid bacteria when it comes to expression systems and genomic engineering[137]. Consequently, the application of *L. lactis* has been expanded from the food industry to being a successful microbial cell factory. Currently, the use of *L. lactis* presents an alternative for membrane protein production[138], protein secretion[139,140] and plant-based protein and secondary metabolite production[141–143]. However, as for *B. subtilis*, there is a need for more genetic tools for strain manipulation.

Table 1: Other microbial cell factories

Organism	Characteristics	Ref.
Saccharomyces cerevisiae	• unicellular eukaryote with long history in brewing and baking • fast growing organism, well-characterized • whole catalogue of engineering tools is available • tolerant to high ethanol concentration and low pH • complex cellular structure that allows for compartmentalization and is able to perform post-translational modifications • production of pharmaceuticals (e.g. insulin, vaccinations and growth hormones), food additives (e.g. vanillin, steviol glycosides) and complex molecules (e.g. polyketides, isoprenoids and opiates	9,144–147
***Aspergillus* spp.**	• filamentous fungi, multicellular eukaryote • tolerant to wide range of temperature and osmolarity • high capacity for secreting homologous and heterologous proteins • is able to perform post-translational modifications • high proteolytic activity • engineering and expression tool are available • production of amylases, lipases and cellulases, therapeutic proteins (e.g. interleukins, antibodies), organic acids (e.g. citric or malic acid), complex compounds (polyketides, isoprenoids), used in bioremediation	148,149

Continuation of Table 1: Other microbial cell factories

***Streptomyces* spp.**	• Gram-positive, filamentous bacteria • GC-rich genome, versatile metabolism • high secretion capacity and low level of extracellular proteases (especially *S. lividans*) • produce most antibiotics of natural origin (e.g. neomycin, tetracycline, chloramphenicol) and other bioactive compounds • limited availability of engineering and expression tool	148,150
Pseudomonas putida	• Gram-negative soil bacterium • versatile metabolism, increased tolerance to oxidative stress and high tolerance to chemical compounds • development of expression tools, such as the SEVA collection, and genetic tools for engineering • used in industrial biocatalysis, bioremediation and bioplastic production	151–157

Chapter 2: Introduction to protein biosynthesis

An understanding of the underlying molecular mechanisms involved in protein biosynthesis is required to address protein production bottlenecks in microbial cell factories. Protein biosynthesis describes the cellular production of new proteins. It comprises central processes in which (i) the DNA is transcribed into messenger RNA (mRNA), (ii) the mRNA is subsequently translated into a polypeptide chain with an amino acid sequence characteristic for each protein, and (iii) the polypeptide chain is folded and transported to the cellular compartment where the protein is supposed to act. Translation, folding and transport can happen simultaneously.

2.1 Transcription in prokaryotes

Transcription is the first step in protein biosynthesis. Here, mRNA is synthetized from a DNA template in three stages: initiation, elongation and termination.

2.1.1 Transcription initiation and regulation

The DNA-dependent RNA polymerase (RNAP) synthesizes mRNA. The bacterial RNAP core enzyme consists of five subunits encoded by *rpoA*, *rpoB* and *rpoC* in *E. coli*. The α subunits bind to the template DNA and serve a regulatory[158] and structural function being chiefly involved in the assembly of the core complex[159]. The β and β' subunit form the catalytic center of RNA synthesis and ensure the binding of the DNA/RNA hybrid formed during transcription. The ω subunit stabilizes the core complex.

A σ factor transiently associates with the core complex and is responsible for recognizing a promoter sequence to orchestrate transcription initiation[160]. In *E. coli*, *rpoD* encodes for the sigma factor σ^{70}, which keeps essential genes and metabolic pathways operating. The σ factor recognizes a consensus promoter sequence that consists of two parts, known as the -10 and -35 sites[161] (Figure 4). The numbers correspond to the approximate distance in nucleotides from the transcription start site (TSS). Compilation of 168 *E. coli* promoter regions found a conserved motive in the -35 (TTGACA) and -10 (TATAAT) regions[162] (Figure 4). Upstream of the promoter core region, promoters can contain an AT-rich sequence, the UP element, which elevates the transcription rate through binding of the α subunits of the RNAP. Together, these characteristics essentially guide the design of synthetic promoters[163].

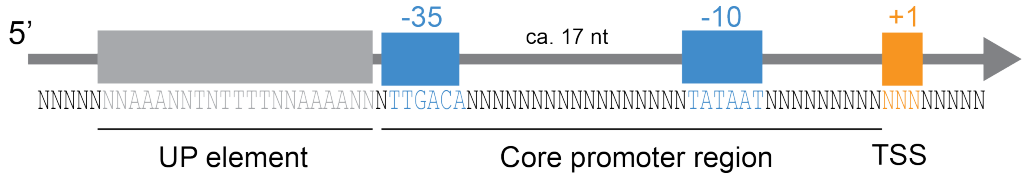


Figure 4: Schematic representation of a prokaryotic promoter. The transcription start site (TSS, +1) is shown in orange. The -10 and -35 sites are shown in blue and located 10 nt and 35 nt upstream of the translational start site, respectively. Promoters can contain an AT-rich UP element (grey). (adapted from Gilman and Love, 2016[163])

Alternative sigma factors get activated during specific stress conditions, such as σ^{38}/σ^{S} (*rpoS*) in starvation and stationary phase[164], which is often found to be mutated in ALE experiments (see Chapter 1)[165,166], σ^{32} (*rpoH*) upon heat exposure which controls chaperones such as GroEL and DnaK and also the Lon and Clp proteases[167] (see section 2.3), or σ^{24}/σ^{E} (*rpoE*) in extreme heat stress and cell envelope stress[168,169]. *E. coli* has a total of six different sigma factors at its disposal together with more than 300 transcription factors (half of which are only predicted) which together fine tune transcription[170].

Having this large number of transcriptional regulators gives *E. coli* the possibility to respond quickly to different environmental conditions. Transcription factors can either turn the transcription of genes on (transcriptional activators) or off (transcriptional repressors) by binding to specific DNA sequences. Transcriptional repressors mostly bind in the promoter region directly interfering with the RNAP, whereas transcriptional activators mostly bind upstream of the promoter to support recruitment of the RNAP[171]. Furthermore, there are transcription factors that can act as both, activators and repressors. Many transcription factors contain a so-called helix-turn-helix motive as DNA binding domain[172]. Most transcription factors need to be activated by e.g. ligand binding or protein-protein interactions. In two component systems, activation is mostly linked to phosphorylation[173]. Transcription factors can play in important role in the development of biosensors in the synthetic biology field (see chapter 3.6).

2.1.2 Transcription elongation

Once transcription is initiated, the RNAP is release from the promoter, which defines the transcriptional speed. In this process, the RNAP creates a so-called transcription bubble. It uses the non-coding strand as template to construct a 5' → 3' mRNA template with a speed of ca. 20 – 40 nt/s[174]. During transcription elongation the RNAP is closely followed by a ribosome immediately translating the synthesized mRNA transcript. While transcription and translation in eukaryotes

take place in different compartments, in prokaryotes transcription rate and translation rate are interconnected in prokaryotes[174]. If translation slows down, the growing gap between polymerase and the ribosome gives access to the Rho factor (see 2.1.3), which triggers termination of the transcription process. Likewise, a faster ribosome can cause secondary structure changes in the mRNA leading to the formation of intrinsic termination hairpins. Conversely, a trailing ribosome can restrain the RNAP from spontaneous backtracking and therefore prevents premature termination[174].

2.1.3 Transcription termination

Once the mRNA transcript is finalized, transcription needs to be terminated. Here, bacteria use two different mechanisms, the Rho-dependent and Rho-independent termination. Rho-independent termination is determined solely by the DNA sequence. These terminator signals contain a ca. 20-nt long GC-rich sequence segment that is followed by a series of adenines. When the GC-rich motif is transcribed, the newly synthesized mRNA transcript will form a hairpin structure that will stall and destabilize the RNAP. Formation of the hairpin structure also destabilizes the RNA-DNA hybrid structure as base pairing between the series of adenines with uracil is rather weak[175,176]. The design of synthetic terminators is based on these principles[177].

Rho-dependent termination requires participation of the Rho factor, a protein that exhibits RNA-DNA-helicase activity. It is deputed whether the protein binds to an unstructured but C-rich sequence, the Rho utilization site, downstream from a stop codon and moves along the RNA towards RNAP, or whether Rho is already associated with the RNAP during elongation. However, it is well established that Rho triggers dissociation of the complex. The *trp* operon in *E. coli* contains one of the best studied Rho-dependent terminators[178,179].

2.2 Translation in prokaryotes

Translation is the second major step in protein biosynthesis. Here, mRNA is translated into a polypeptide chain in three phases: initiation, elongation and termination. Translation initiation occurs as soon as the 5' end of a mRNA strand is synthesized. The heart of the translation process is the ribosome, which consists of the two ribosomal subunits 30S and 50S, together forming the functional 70S ribosome. Both ribosomal subunits are constructed of ribosomal RNA (rRNA) and proteins (62% RNA, 38% protein)[180]. An mRNA can be occupied with several ribosomes, forming

a so-called polysome. Both, translation initiation and translation elongation contribute to the frequency of ribosome loading in polysomes[181,182].

2.2.1 Translation initiation

As with transcription, the initiation of translation is rate-limiting. Three initiation factors IF1, IF2 and IF3 are involved in the assembly of the initiation complex. IF1and IF3 prevent the premature congregation of the large to the small ribosomal subunit[183]. IF3 also aids the binding of the new mRNA[184]. The selection of an initiation start site depends on the interaction of mRNA and the 30S subunit. The mRNA contains a so-called Shine-Dalgarno (SD) sequence located circa 4 to 14 nucleotides upstream of the start codon that is recognized by the 30S subunit which contains the complementary sequence in the 16S rRNA[185] (Figure 5). The 3'-terminal sequence of *E. coli* 16S rRNA was shown to be 3'-AUUCCUCCA-. Thus, the consensus core SD sequence in *E. coli* is 5'-AGGAGG-3'. Together, the two sequences form a double stranded RNA structure[186]. Due to this mechanism, proteins with a non-AUG start codon, such as those of the *lac* operon, can be translated[33,187,188]. Experiences from protein production suggest that the SD sequence and its surrounding sequence play a very important role in the efficiency of translation initiation[189].

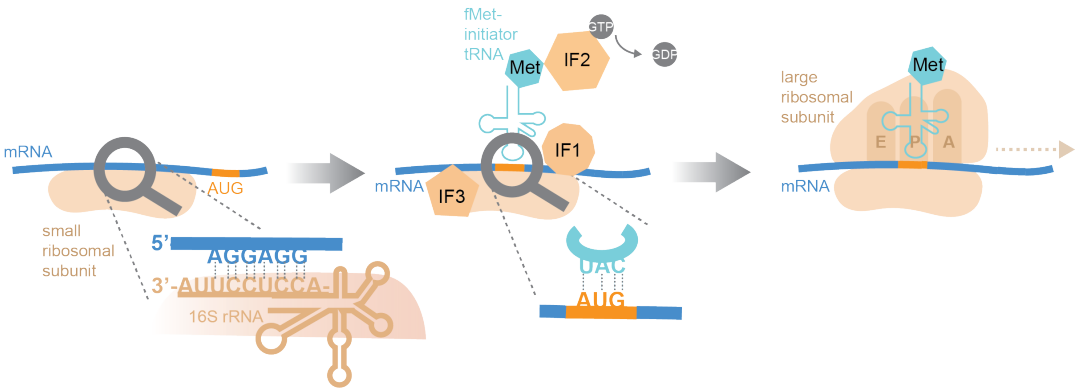


Figure 5: Translation initiation in prokaryotes. The 3'-end of the 16S rRNA binds to the Shine-Dalgarno sequence. IF1 and IF3 prevent premature assembly of the ribosome. IF2 guides the initiator tRNA to the AUG start codon. Upon GTP hydrolysis the initiation factors leave the small subunit and allow the ribosome to assemble and start translation.

IF2 binds specifically to the N-formylmethionine-tRNA (fMet-tRNA), also called initiator tRNA, and guides it to the ribosome. Thus, independent of the codon utilized as translational start site, all proteins start with a methionine. The 30S subunit, mRNA and initiator tRNA form a preinitiation complex[186,190]. IF1, IF2 and IF3 dissociate from the complex after hydrolysis of GTP, which allows the 50S subunit to bind to the 30S subunit and to complete the initiation complex.

How to optimize and engineer the translation initiation region for most efficient protein production is discussed in more detail in Chapter 3.4.4. The optimization of translation initiation is subject of two papers in this thesis and led to optimized protein production in Gram-negative and Gram-positive hosts (Paper 1 and 3).

2.2.2 Translation elongation

tRNAs direct the correct amino acid to the given position in the mRNA. Each tRNA contains an exposed triplet of nucleotides (anti-codon) that is complementary to the codons of the mRNA[191]. Anti-codon and codon form hydrogen bonds that maximize stability when paired correctly. The aminoacyl-tRNA-synthetase catalyses the loading of the tRNA. Under ATP hydrolysis the amino acid is attached to the C-terminal adenosine of the acceptor arm[191,192]. A loaded tRNA is called aminoacyl-tRNA. The ribosome contains three active sites, A-, P- and E-site. The A-site is the point of entry where each aminoacyl-tRNA arrives. The P-site contains the elongating polypeptide chain, bound to a tRNA. Uncharged tRNAs leave the ribosome through the E-site[193] (Figure 6).

Three elongation factors (EF) orchestrate the elongation process. EF-Tu, a small GTPase, facilitates the binding of an aminoacyl-tRNA to the A-site. If the correct aminoacyl-tRNA is in place, GTP hydrolysis by EF-Tu will lead to a large but energetically favourable conformational change in the ribosome, which is called conformational proofreading[194,195]. The elongation factor EF-Ts resets EF-Tu by exchanging GDP with GTP. Once an aminoacyl-tRNA is bound in the A-site, the polypeptide chain is removed from the tRNA in the P-site and its carboxyl group is bound to the amino group of the newly arrived amino acid, a process known as peptide bond formation, catalysed by the 23S rRNA of the large ribosomal subunit[193,196] (Figure 6). For a moment the polypeptide chain is then associated with the tRNA at the A-site, but shortly after the translocation of the ribosome relative to the mRNA follows, which is catalysed by elongation factor EF-G[197] (Figure 6).

The A-site can then be occupied by the next aminoacyl-tRNA assisted by EF-Tu. The binding of a new aminoacyl-tRNA leads to the release of the deacylated tRNA in the E-site because of an allosteric linkage between the A- and E-site[198]. This process continues until the ribosome reaches a stop codon. The large ribosomal subunit contains a polypeptide exit tunnel through which the growing polypeptide chain will leave the ribosome[199]. The ribosome is able to elongate the polypeptide chain by 10 to 20 amino acids per second[200]. Folded mRNA structures slow down elongation, yet the ribosome exhibits helicase activity and is able to quickly unwind these

structures[201,202]. The ribosome's helicase activity was a decisive factor in the design process of the RNA hairpin structures, which are utilized throughout the publications included in this thesis (Paper 1-3).

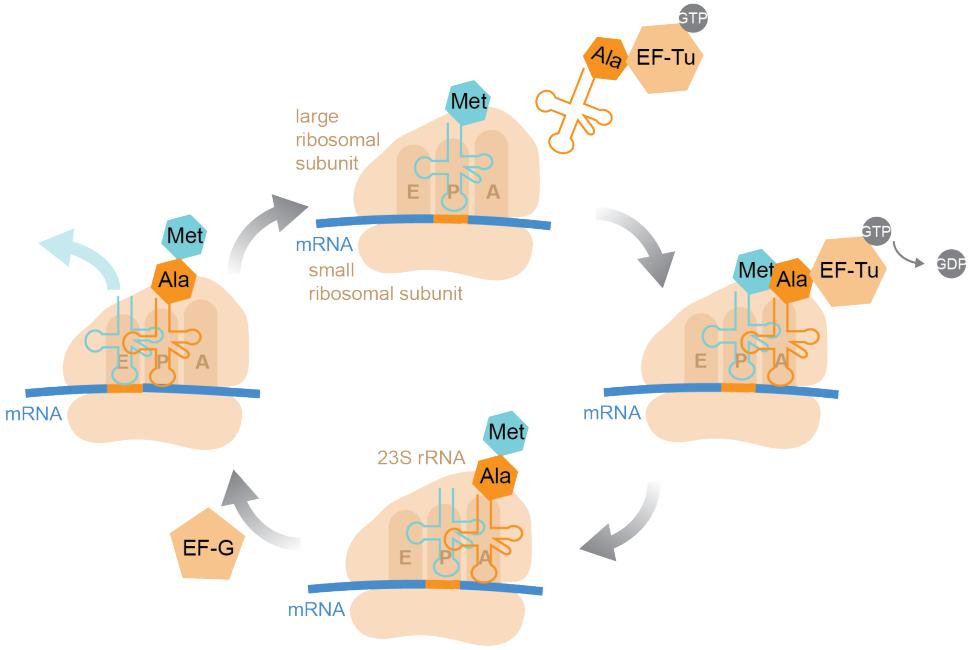


Figure 6: Translation elongation in prokaryotes. After the fMet-tRNA is bound in the P-site, an aminoacyl-tRNA can bind to the A-site. EF-Tu facilitates binding and GTP hydrolysis leads to a conformational change. Catalysed by the 23S rRNA, a new peptide bond is formed and the polypeptide chain is then associated with the tRNA at the A-site and a deacylated tRNA is located at the P-site. The following translocation of the ribosome, catalysed by EF-G, translocates the polypeptidyl-tRNA to the P-site and the deacylated tRNA to the E-site. The binding of a new aminoacyl-tRNA to the A-site, releases the deacylated tRNA at the E-site. This iterative cycle is repeated until the stop codon is reached.

2.2.3 Translation termination and ribosome recycling

A soon as the ribosome reaches a stop codon the termination of the translation process begins. This process is orchestrated by three ribosome release factors and assisted by a ribosome recycling factor (RRF). Release factor RF1 recognizes all stop codons whereas RF2 recognizes UAA and UAG[193,203]. Both factors trigger the release of the polypeptide chain. RF3 catalyses the release of RF1 and RF2 by GTP hydrolysis[193]. After the release of the polypeptide chain the ribosome is left with the deacylated tRNA and the mRNA. RRF triggers the ribosomal dissociation into large and small subunit together with the elongation factor EF-G[193,203]. Ribosome disassembly requires the hydrolysis of GTP by EF-G[204]. Subsequently, IF3 is required to replace the deacylated tRNA in the P-site, which releases the mRNA. All components are free for a new translation initiation[183,205]. In operons, ribosome re-initiation can occur shortly after termination. Operons are genes that are transcribed into one mRNA but are translated individually. Stop and start codons of the coding

sequences are in close proximity, which enables the upstream elongating ribosome to dissociate and reassemble at the downstream intergenic region, and thereupon initiate translation of the downstream coding sequence. However, due to the ribosomes helicase activity, the mRNA structure around the translation initiation region of the downstream gene will be in a relaxed state when the elongating ribosome reaches the end of the upstream gene. This is favourable for ribosome binding and assembly, whereupon *de novo* initiation of the downstream coding sequence can occur[206]. Efficient translation initiation of the first gene in an operon will lead to accelerated translation of the following genes (see Appendix 1).

2.2.4 Protein synthesis inhibitors

Besides intrinsic regulatory processes controlled by enzymes, small molecules or peptides can inhibit protein synthesis. Most inhibitors of protein synthesis are antibiotics and interfere at different stages of translation. The antibiotic neomycin prevents the assembly of the ribosome by binding to the 30S subunit[207]. Tetracyclines block the A-site and therefore prevent the binding of aminoacyl-tRNAs[208,209]. Many of the aminoglycosides, such as kanamycin and gentamycin, and the aminoglycoside-like spectinomycin intervene in ribosomal translocation and can lead to mistranslation by tRNA mismatching[210,211]. Chloramphenicol blocks the polypeptidyl transfer reaction of the 50S subunit[212]. Macrolides, such as erythromycin, can bind as well to the 50S subunit and therefore prevent the peptidyl transfer reaction[212]. Other macrolides inhibit ribosome translocation or cause premature dissociation of the peptidyl-tRNA from the ribosome. Antibiotics such as rifamycin or rifampicin are an exception and inhibit transcription by binding to the β-subunit of the core complex of the RNAP[213].

In a protein production project one should therefore carefully evaluate what antibiotic resistance gene is best to use. The corresponding antibiotic that needs to be supplemented for plasmid selection can interfere with protein production. Antibiotic resistance genes are discussed in greater detail in Chapter 3.1.2.

2.3 Protein folding and targeting

Translation transforms genetic information into amino acid sequences but does not produce a functional protein. To this end, the polypeptide chain needs to fold into a specific three-

dimensional conformation. Folding is a spontaneous process that is determined by the formation of intramolecular hydrogen bonds, hydrophobic interactions and electrostatic interactions and reaches a final native conformation which has the lowest energetic state[214]. If in a non-native conformation, hydrophobic residues are often exposed to the aqueous environment and cause aggregation of the protein[215]. To counteract this effect, many proteins require assistance by molecular chaperones to reach their folded states fast and efficiently[216]. Most chaperones are proteins but in some rare cases RNA and lipids can also influence the folding of a protein[217,218]. While a detailed review of protein folding and targeting is well beyond the scope of this thesis, a short summary of the chaperones and translocation systems of *E. coli* is given.

2.3.1 Protein folding in the cytoplasm

Three chaperones, the trigger factor (TF), DnaKJ/GrpE (Hsp70 system) and GroEL/ES (Hsp60 system) are the central elements of protein folding and quality control in the cytoplasm of *E. coli*[219,220]. They work as an interactive network passing unfolded proteins from one to the other[221] (Figure 7).

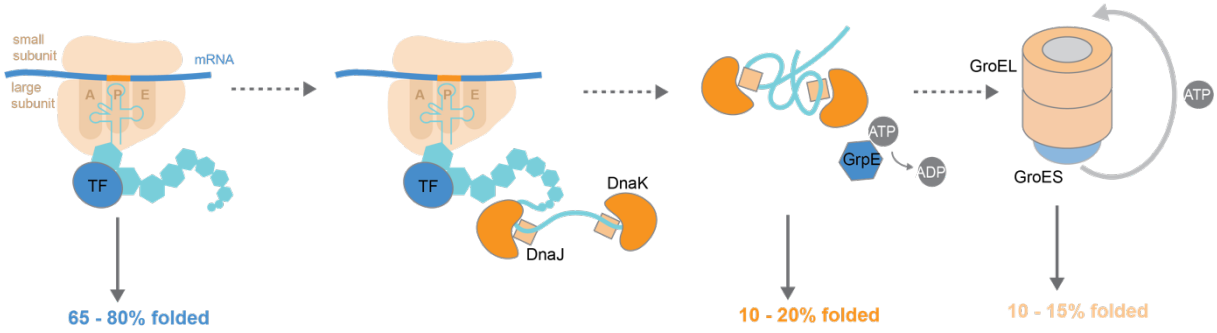


Figure 7: Cytoplasmic chaperones. The trigger factor (TF) is a ribosome-associated ATP-independent chaperone that directly binds to the polypeptide chain leaving the ribosome's exit tunnel. Upon interaction with TF 65 to 80% of total proteins are folded into their native state. DnaJK receive substrates from TF and fold proteins upon ATP hydrolysis by GrpE. 10 to 20% of the total proteins are folded into their native state by the DnaJK-GrpE chaperone system. The last 10 to 15% are passed on to the GroEL/GroES complex where are folded in an ATP-dependent manner. (adapted from Kim *et al.*[219]).

The trigger factor is a ribosome-associated chaperone whose role it is to prevent misfolding of polypeptide chains leaving the ribosome's exit tunnel[222] (Figure 7). DnaKJ/GrpE, receives substrates from TF[223] (Figure 7). It assists folding in a co- and post-translational manner in an ATP-dependent reaction[219,221,223]. Unfolded proteins will be passed on to the GroEL/ES chaperonin for further folding assistance (Figure 7)[219]. The GroEL/ES chaperonin is required under all growth conditions indicating that also essential genes require GroEL/ES folding assistance[224,225]. Due to its distinct architecture (a barrel structure with a lid)[226,227], a fully assembled GroEL/ES complex

provides protection from the cellular environment[225]. If the substrate is still not correctly folded after release from the chaperonin, it will rebind to GroEL/ES and refold[225].

2.3.2 Co-translational translocation and insertion of inner membrane proteins

The Gram-negative cell envelope contains two membranes, which separate the cell from its environment. These biological membranes are dynamic assemblies of lipids and proteins. The inner membrane forms a symmetric lipid bilayer consisting of phospholipids and contains integral or peripheral membrane proteins. Inner membrane proteins (IMPs) are α-helical proteins. The outer membrane is composed of an asymmetric lipid bilayer, since the outer leaflet contains glycolipids instead of phospholipids, which mainly serve as a barrier towards external threats. The majority of outer membrane proteins (OMPs) are β-barrel proteins. The space between the two membranes constitutes the periplasm, which, in contrast to the cytoplasm, is an oxidizing environment[228].

Most IMPs are inserted into the membrane during translation. This process is called co-translational insertion and is mostly mediated by the Sec-translocon, a protein-conducting channel, which consists of three integral membrane proteins, SecY, SecE and SecG, and the peripheral associated ATPase SecA[229–231]. Co-translational translocation or insertion is mediated via the SRP (signal recognition particle) pathway[232] (Figure 8). SRP prevents folding before the inner membrane is reached and translocation is started[232,233]. The SRP receptor FtsY mediates the transfer of the polypeptide chain chain to the Sec-translocon[234–236]. The inner membrane protein YidC assists the IMP integration process in many cases[237]. YidC can also promote membrane protein insertion itself by functioning as a membrane insertase[220,238–240].

Whether a polypeptide chain emerging from the ribosome is targeted to the inner membrane depends on an N-terminal signal or anchor sequence[241]. Signal peptides typically have an average size of 20 amino acids with a tripartite structure: a positively charged N-terminal region (n-region), a hydrophobic core (h-region) and a polar C-terminal region (c-region). Signal sequence hydrophobicity indicates if a protein is translocated co- or post-translationally[229,242,243]. Signal sequences of post-translationally translocated proteins are less hydrophobic than those targeted by SRP.

2.3.3 Post-translational translocation

Most periplasmic and OMPs, are translocated post-translationally[244]. In these cases, TF competes with SRP for binding to the emerging polypeptide chain[244,245]. Upon binding of TF post-translationally translocation depends on the cytoplasmic chaperone SecB, which prevents premature folding and mediates targeting to the Sec-translocon[230] (Figure 8). SecB binds to SecA associated to the protein conducting channel[246].

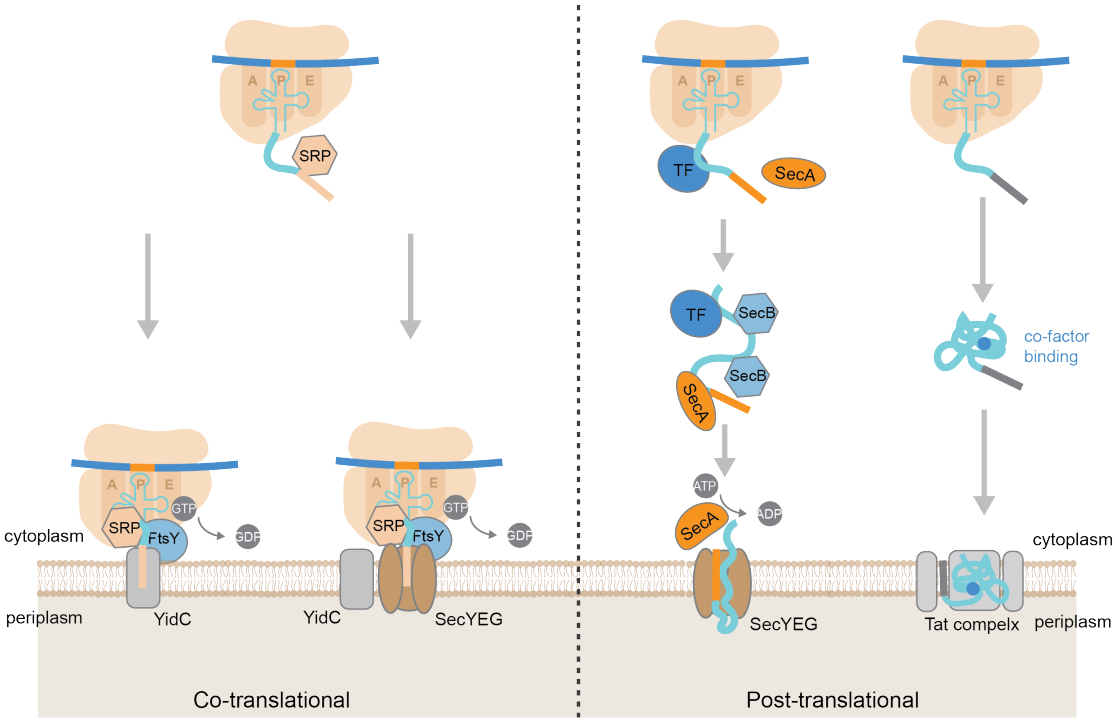


Figure 8: **Co- and post-translational transport across the inner membrane.** Co-translational protein insertion or translocation occurs *via* the SRP-dependent pathway through the SecYEG translocon or the YidC protein. Post-translational translocation occurs *via* SecA/B-dependent translocation in an unfolded state through the Sec-translocon or in a folded state *via* the Tat complex.

Besides being targeted to the Sec-translocon, secretory proteins can also be translocated *via* the twin-arginine translocation system (Tat system). The Sec-translocon transports unfolded proteins across the inner membrane whereas the Tat system serves to transport folded proteins[230,247]. Tat-targeted proteins may bind with cofactors before translocation, an important feature for many redox pathways[248,249] (Figure 8). Tat substrates use the cytosolic chaperones TF, DnaKJ/GrpE and GroEL/ES for folding[250]. The overall structure of Tat signal peptides is similar to those of Sec signal peptides[241]. However, Tat signal peptides are larger than Sec signal peptides due to an extension in the N-terminal region by the consensus motive SRRxFLK, which contains two conserved and almost invariant arginine residues[251,252]. Compared to the Sec-translocon, transport *via* the TAT pathway is slower and less efficient[252,253].

2.3.4 Protein folding in the periplasm

Like proteins in the cytoplasm also periplasmic proteins require assistance with folding. YfgM is a recently discovered inner membrane protein, which traffics of proteins between the Sec-translocon and the periplasmic chaperone machineries DegP, Skp, SurA, FkpA, Spy and PpiD[254,255] (Figure 9). DegP, Skp and SurA are involved in OMP translocation and folding and will be discussed further in section 2.3.5.

The inner membrane chaperone PpiD supports folding of newly translocated polypeptides, including periplasmic proteins and OMPs[256]. Spy is a recently discovered periplasmic chaperone, whose function is unknown. It was first believed to interact only with periplasm-destined proteins by binding to aggregation-prone regions[257] but more recent studies show that it can rescue the loss of FkpA and Skp, two chaperones involved in guiding OMPs to the β-barrel assembly machinery[258]. FkpA is another periplasmic chaperone with *cis*/*trans* peptidyl-prolyl isomerase (PPIase) activity that has been shown to mediate the folding of periplasmic and outer membrane proteins[259].

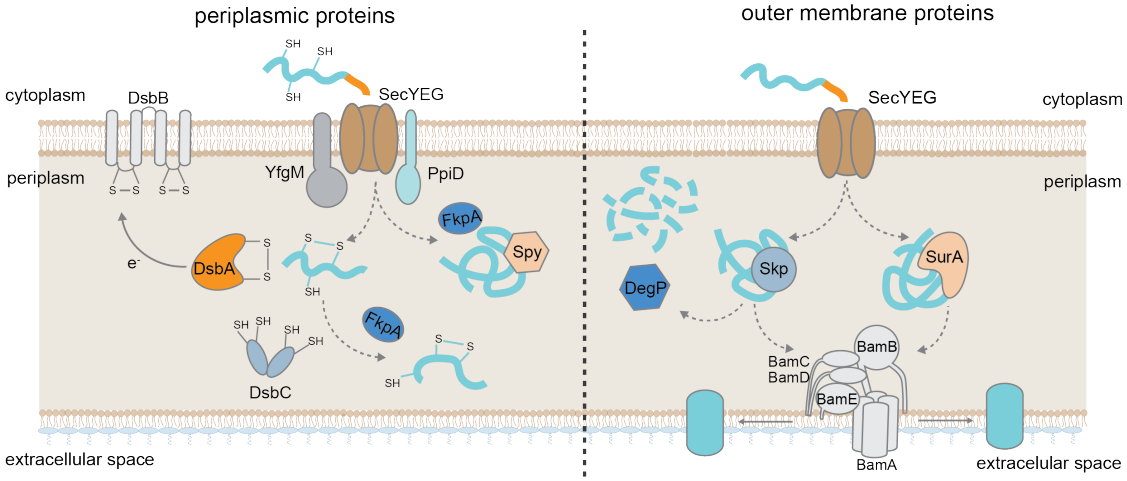


Figure 9: **Folding and targeting in the periplasm.** YfgM trafficks proteins to the periplasmic chaperones. PpiD supports folding of newly translocated peptides. Disulfide bonds are inserted by DsbA and isomerised by DsbC. DsbB regenerates DsbA. Spy is a chaperone with unknown function. FkpA is a peptidyl-prolyl isomerase. SurA and Skp are involved in the folding and trafficking of outer membrane proteins to the BAM complex. DegP is a periplasmic protease involved in the degradation of misfolded outer membrane proteins.

In *E. coli* about 300 proteins are estimated to contain at least one disulfide bond[260]. Disulfide bonds can be important for the stability and function of a protein and their formation is only stable in the oxidizing environment of the periplasm[261]. The soluble oxidoreductase DsbA generates disulfide bonds (Figure 9). DsbA is then regenerated by DsbB. DsbB is an IMP which transports the electrons to the Q-cycle of the respiratory chain[220,261]. A third component in the system is DsbC, a disulfide bond isomerase[262,263].

2.3.5 Folding and insertion of outer membrane proteins

The chaperones SurA, Skp and DegP are involved in the trafficking of OMPs through the periplasm to prevent misfolding and keep the peptide chain partly unfolded until the β-barrel assembly machinery (BAM) complex is reached[264] (Figure 9).

The PPIase SurA is the main periplasmic chaperone in *E. coli* involved in OMP biosynthesis[265,266]. It is the only chaperone that has been shown to interact with the BAM complex[267]. SurA is also involved in targeting Tat substrates to the outer membrane[268]. Skp is a holdase chaperone, which binds hydrophobic patches of substrates with its hydrophobic cavity, and keeps them from folding, while simultaneously allowing hydrophilic patches to fold. The OMP OmpA is the most prominent substrate of Skp[269,270]. It is believed that Skp can target the BAM complex but experimental evidence is not available[220]. DegP was previously believed to show chaperone activity in the periplasm, but more recent discoveries suggest that DegP mainly functions as a protease and degrades misfolded OMPs which cannot be rescued[168,271]. The BAM complex, consisting of five subunits, assists the insertion of OMPs[272,273] (Figure 9). Several structures of the BAM complex and mechanisms for OMP insertion have been described[274–276]. However, the exact mechanism remains elusive.
Under normal conditions *E. coli* only secretes very few proteins into the extracellular space. A review by Costa *et al*. gives a comprehensive overview of all secretory systems[277].

Chapter 3: Recombinant protein production in bacterial cell factories

Equipped with a comprehensive overview and understanding of protein biosynthesis one can then proceed to design, build and test, and optimize, recombinant protein production in microbial cell factories. The emergence of recombinant protein production has revolutionized scientific research and industrial biotechnology. The possibility to produce proteins in suitable host organisms enabled characterization, functional and structural studies of many unknown proteins and brought many therapeutic proteins into clinical application[10]. Moreover, recombinant protein production allows for engineering of proteins with improved characteristics such as solubility, activity or affinity and enables tagging of proteins for detection and purification[129,278]. As more explicitly discussed in Chapter 1.1, the emergence of synthetic biology impacted the rapid development of recombinant protein production.

Despite all advances, however, bottlenecks in recombinant protein production are encountered regularly. Production of insoluble aggregated forms, degradation, low yields and toxicity limit the success of recombinant protein production. This chapter focusses mainly on overcoming bottlenecks and optimizing recombinant protein production in *Escherichia coli* by engineering gene, vector and strain (Figure 10), but also gives a brief insight into the increasingly popular hosts *Lactococcus lactis* and *Bacillus subtilis*. The choice of the production host, some of which are reviewed in more detail in section 1.3., is the first and very important step in a protein production project.

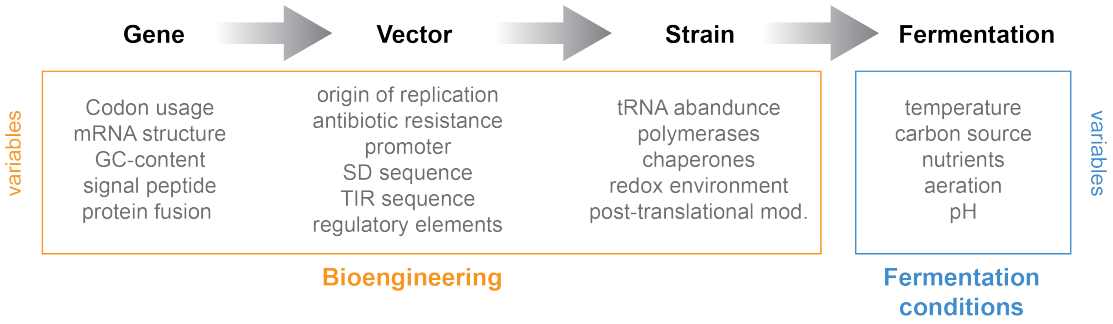


Figure 10: Optimization variables for protein production. Engineering at gene, vector and strain level can benefit protein production. Finally, it is crucial to define optimal fermentation conditions for large-scale protein production for industrial or pharmaceutical applications. (adapted from Gustafsson *et al*. 2012[97]).

3.1 Bacterial expression vectors and their design

Besides the choice of expression host, the choice of an expression vector is a decisive factor for successful protein production. In addition to the multiple cloning site, where the gene of interest is cloned into, an expression vector consists of three major parts – the promoter, driving expression of the target gene, the origin of replication, determining the copy number and the antibiotic resistance marker, needed to maintain the plasmid in the cell (Figure 11).

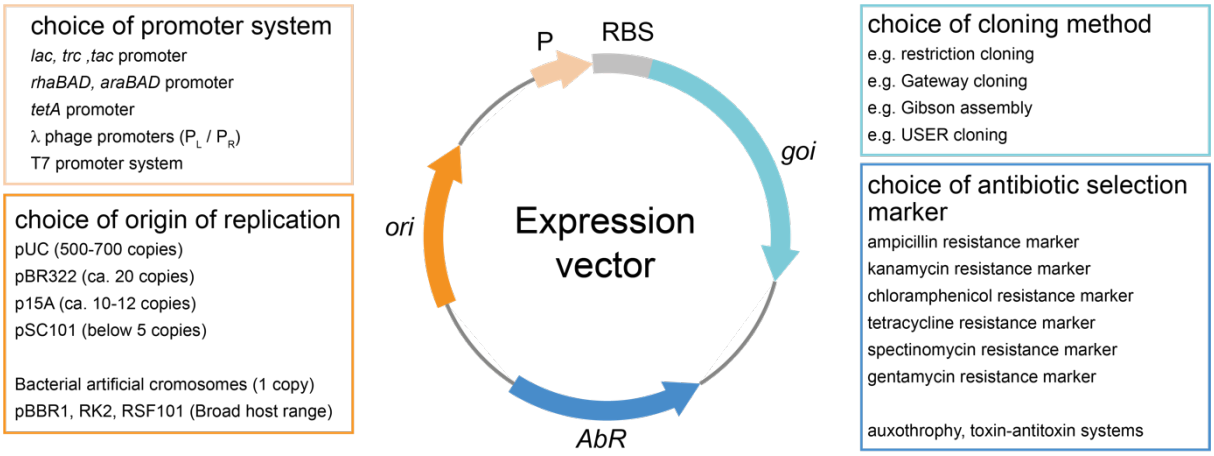


Figure 11: General structure of a prokaryotic expression vector. A promoter (P) drives the expression of a gene of interest (goi), which can be cloned into the vector with different cloning techniques. The ribosome binding site (RBS) is a decisive factor for translation efficiency. The antibiotic resistance marker (AbR) is used for selection and maintenance of the plasmid. The origin of replication (ori) determines the copy number of a vector. Examples for promoter, origin of replication, antibiotic selection marker and cloning methods are listed in the boxes.

3.1.1 Promoter systems

Apart from the in section 2.1. mentioned RNAP binding motives, located 10 and 35 nucleotides upstream of the transcription start site[162], strength and regulation of a promoter depend on distinct regulatory elements[71] that react to specific factors, such as small molecules or increased temperature[279–282]. Promoters in an expression vector can either be derived from the expression host, e.g. from operons involved in sugar utilization, or can be rationally, synthetically designed[279–281,283]. Here, the promoters used in the studies of this thesis are described in more detail. Other promoter systems that are frequently used in protein production setups are listed in Table 2.

The rhaBAD promoter

The *rhaBAD* promoter is originally driving the expression of the L-rhamnose metabolizing pathway RhaBAD. Apart from that, the rhamnose utilization system encodes the rhamnose transporter RhaT and the regulatory proteins RhaS and RhaR (Figure 12). All three promoters are

activated upon addition of L-rhamnose. In the presence of L-rhamnose RhaR activates transcription of *rhaSR*. Production of RhaS leads to the activation of the *rhaT* and *rhaBAD* promoter[284–286]. In a protein production set-up the *rhaBAD* promoter drives expression of the gene of interest. The regulators RhaR and RhaS are normally co-expressed from the expression plasmid (Figure 12). The rhamnose inducible promoter system is extremely tight and highly tunable, which makes it very attractive for the expression of toxic proteins[287]. In this thesis, the tunability of the P_{rhaBAD} promoter system was exploited to invest the dynamic range of several commonly used antibiotic resistance markers and to validate the translational coupling devices that were developed in the course of Paper 1.

As so often for sugar-inducible promoter systems, P_{rhaBAD} is susceptible to carbon catabolite repression[284,288]. Carbon catabolite repression enables bacteria to utilize sugars in a sequential order[289]. The cyclic AMP receptor protein (CRP) binds upstream of the polymerase binding site to a so called CRP binding site[290] functioning as a transcriptional activator. CRP is activated upon binding of cyclic AMP (cAMP). cAMP levels are elevated if no glucose is available and *vice versa*[289].

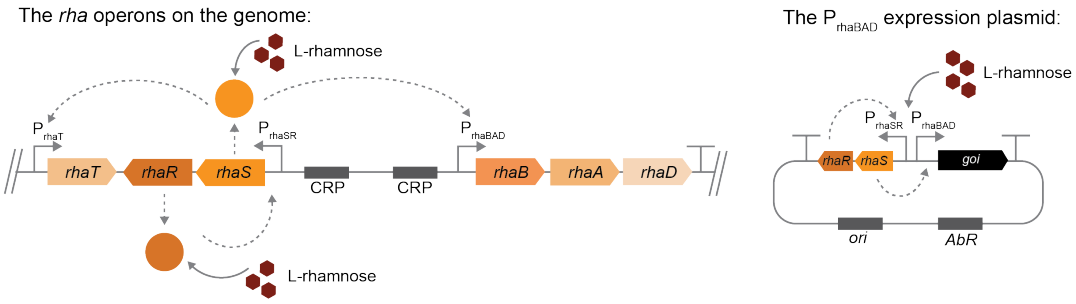


Figure 12: *The rhaBAD* promoter system. In the presence of L-rhamnose RhaS activates P_{rhaBAD}, which encodes the enzymes RhaBAD for rhamnose untilization, and P_{rhaT}, which encodes the rhamnose transporter RhaT. RhaR activates P_{rhaSR}, which encodes the two regulators RhaR and RhaS and, thus amplifies the inducing effect. (inspired by Hjelm[220])

In general, sugar-inducible promoters have the drawback that the cell can metabolize the inducer over time. Therefore, inducers need to be supplied regularly in large-scale fermentation processes. A recent study demonstrated that the tunability of the *rhaBAD* promoter is caused by consumption of L-rhamnose rather than by regulation of protein production rates. Inactivation of the rhamnose-utilizing enzyme RhaB led to these findings. An additional knockout of the rhamnose transporter RhaT restored a tunable phenotype and varying the concentration of L-rhamnose could precisely set protein production rates and the system was stable over time[291]. Another study used the L-rhamnose analogue L-mannose, which cannot be catabolized by *E. coli*, for more controlled protein

production by P_{rhaBAD}. L-mannose can only be used for induction when simultaneously the regulator RhaS is constitutively over-expressed[292].

The T7 promoter

Despite the necessity of the bacteriophage T7 RNAP being encoded in the host cell, either on the genome or supplemented on a plasmid, the T7 expression system is one of the most popular expression systems for protein production since it was introduced in the 1980s. The T7 RNAP is highly active and very selective for its own promoter, which does not naturally occur in *E. coli*[293]. *E. coli* BL21(DE3) is the most common strain used for T7-based expression. In addition to carrying the bacteriophage DE3 gene encoding for the T7 RNAP on its genome, BL21(DE3) lacks the *lon* protease and is deficient in the *ompT* outer membrane protease which degrade proteins during purification[294].

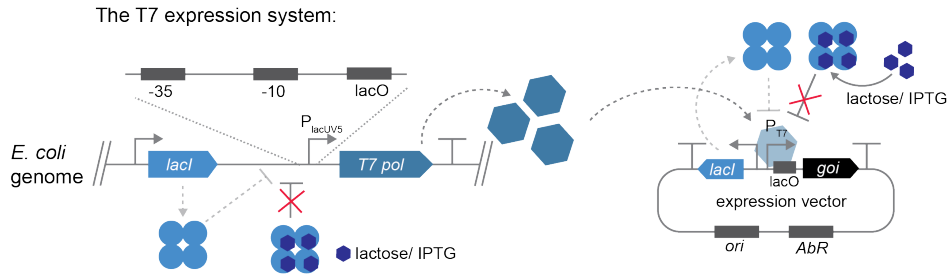


Figure 13: The T7 expression system. The T7 RNAP is under control of P_{lacUV5}, which is recognized by the endogenous RNAP of *E. coli*. T7 RNAP recognizes only the T7 promoter, which controls expression of the gene of interest (goi). (inspired by Hjelm [220]).

The T7 RNAP is controlled by a strong *lac* derived *lac*UV5 promoter, which itself is recognized by endogenous *E. coli* RNAP[295]. The *lac*UV5 promoter is commonly induced with the metabolically inactive lactose analogue isopropyl β-D-1- thiogalactopyranoside (IPTG). In the absence of IPTG, the repressor LacI occupies the operator site (lacO) and prevents expression of the T7 RNAP (Figure 13). LacI is only removed when the binding of IPTG induces a conformational change of the repressor. For recombinant protein production a T7-*lac* hybrid promoter, which also contains the LacO operator site, drives the expression of the target gene. Thus, LacI represses both, expression of the T7 RNAP and expression of the target gene[296] (Figure 13). Due to its high transcription rate, T7 RNAP produces large amounts of mRNA resulting in large amounts of protein[297]. However, this turns out to be problematic for membrane or secreted proteins, which then encounter saturation of the secretion machinery. Reduction of mRNA levels can be achieved by expressing a gene encoding the T7 lysozyme, a natural inhibitor of the T7 RNAP. The T7 lysozyme is encoded on the plasmids pLysS and pLysE providing low and high levels of T7

lysozyme, respectively. As a consequences of this, co-expression of T7 lysozyme also prevents leaky expression of the target gene[298,299]. Throughout this thesis, the T7 expression system is used for protein production in *E. coli* BL21(DE3) (Paper 1, Paper 2 and Paper 4).

Table 2: Commonly used promoter systems for protein production in *E. coli*.

Promoter	Inducer	Characteristics	Ref.
P_{lac}	lactose, IPTG	• Best studied promoter system • Weak promoter, can be used for low level production of toxic proteins or membrane proteins • repressed by LacI, which binds to a specific operator site in the promoter region (lacO) and is only removed when the binding of lactose induces a conformational • is controlled by carbon catabolite repression	25,289,290,300
P_{tac}, P_{trc}	lactose, IPTG	• hybrids of the tryptophan (*trp*) promoter and P_{lac} • vary in their spacing between the -10 and -35 region • three times stronger than the *trp* promoter and ten times stronger than the *lac* promoter • repressed by LacI	301,302
P_{araBAD}	L-arabinose	• tight promoter, limited titratability, as increasing the arabinose concentrations mostly shortens the delay time until expression but does not substantially increase production • repressed by AraC in the absence of arabinose, activated by AraC in the presence of arabinose • pBAD vectors are a commercial vector series with the *araBAD* promoter • can lead to mixed expression populations, constitutive expression of the transporter AraE results in homogenous cultures • is controlled by carbon catabolite repression	280,303–307
P_{tetA}	tetracycline, anhydro-tetracycline (aTc)	• tight, strain independent promoter • repressed by TetR in the absence of tetracycline • self-repressing system, which makes it capable of fast and efficient response to tetracycline and very attractive for recombinant protein production • independent of carbon catabolite repression • use of an antibiotic as inducer can lead to growth interference problems, use of the non-toxic tetracycline derivative anhydrotetracycline (aTc) recommended	308–312
P_L, P_R	Temperatures above 37°C	• λ phage promoter systems • regulated by the heat-sensitive cI repressor • great alternatives in cases where the addition of a chemical compound as inducer is not desire	282,313

3.1.2 Selection marker

Transforming a vector into bacterial cells is an inefficient process. In order to get and maintain a vector in the cell, an expression vector needs to encode a selection marker. Generally antibiotic resistance genes are used as selection markers in bacteria. By adding the respective antibiotic to the culture one makes sure that only the cells carrying the expression vector will survive. Without selection pressure, cells not containing the vector will have growth advantages and overgrow the plasmid carriers.

The most common antibiotic resistance markers in *E. coli* confer resistance to ampicillin, kanamycin and chloramphenicol; spectinomycin, tetracycline and gentamycin are also used frequently. Resistance to kanamycin, chloramphenicol and spectinomycin is conferred by three cytosolic transferases. Ampicillin acts in the periplasm, which is the reason why the ampicillin resistance protein β-lactamase likewise is secreted into the periplasm where it degrades the β-lactam antibiotic. It is therefore possible that β-lactamase competes with membrane and secretory proteins for the secretion machinery. In a study conducted in the course of Paper 1, we saw elevated membrane protein expression levels, when the co-expressed ampicillin resistance marker was replaced with a spectinomycin resistance marker (see Paper 1, Supplementary Figure 2). Tetracycline resistance is transferred by the efflux pump TetA which catalyzes an energy-dependent export of tetracycline, causing similar competition for secretion resources[314]. As already mentioned in section 2.2.4 most antibiotics act on the ribosome and might therefore interfere with protein biosynthesis. In a study by Kim *et al*., included as a publication in this thesis (Paper 4), 122-fold differences in protein production were found by just exchanging the antibiotic resistance marker[315]. These production differences most likely arise from the fitness costs, the metabolic burden and physiological changes of the cell caused by antibiotic selection[316–318].

Problems of using antibiotics in large-scale recombinant protein production have long been recognized, but only few alternative methods for selections have been developed. Here, essential genes and toxin-antitoxin systems have been proven to be alternative methods resulting in very stable plasmid system[319,320]. Both approaches come with the disadvantage that they rely on a genomic modification in the expression strain prior to its use for protein production.

3.1.3 Origin of replication

Different expression vectors are maintained at different copy numbers in the cell depending on their origin of replication. Vectors are divided into high, medium and low copy plasmids. For cloning and DNA amplification purposes high copy plasmids, such as ColE1 based replicons, are preferable. The native ColE1 origin can be found in 20 to 30 copies in a cell[321,322] but mutants such as in the pUC vector series produce between 500 and 700 copies per cell[323–326]. The frequently used pBluescript plasmid is derived from the pUC vector series[327]. A high copy number is not always desirable and can cause problems when producing toxic proteins and membrane proteins in particular. In the case of a high copy plasmid, the slightest leakiness of the promoter will already lead to protein accumulation. In these cases, medium copy number plasmids, such as the pBR322 and p15A replicon can be used. The pBR322 replicon originates from the ColE1-like replicon pMB1 with a copy number of 15 to 20[328,329]. The p15A origin can be found in the pACYC vector series and has around 10 to 12 copies per cell[129,278,330]. Plasmids with the pSC101 origin count as low copy plasmids with less than 5 copies per cell[26,331]. The pSC101-derived plasmid pSC101-ts has a point mutation in an essential replication protein, which makes replication temperature- sensitive[332]. Plasmids with similar origins of replication are unable to coexist in the same cell. This phenomenon is called plasmid incompatibility. If two plasmids require the same functions and machineries for replication upon partitioning into daughter cells one plasmid can be lost over time[278].

The ColE1, p15A and pSC101 replicons only function in *E. coli* and some other *Enterobacteriaceae*[333]. If working in another organism than *E. coli,* one can exploit vectors for broad host range, such as pBBR1[334,335], RK2[336], RSF1010[337] or pRO1600[338]. Many of those broad host range vectors are shuttle vectors that can also replicate in *E. coli*. Often plasmid manipulations can be accomplished much easier in *E. coli* and the vector is subsequently transferred to the host organism[339]. Vectors encompassing the R6K replicon are often used as suicide vectors[155]. The replication mechanism of R6K*oriV* requires the π protein for initiation of replication, which is not supplemented in a suicide vector[340].

In a study by Kim *et al*., included as a publication in this thesis (Paper 4), 49-fold differences in protein production were found by just exchanging the origin of replication[315].

3.1.4 Gene expression systems for *Lactococcus lactis*

The comprehensive knowledge about its genome[341] does not only encourage general interest in *Lactococcus lactis*, but greatly simplifies the development of gene expression systems. Various constitutive and inducible expression systems have been developed in the last years. Available promoter systems are listed in Table 3. High-level production of heterologous proteins has been achieved using constitutive or inducible promoters. Nevertheless, inducible promoters are generally desired for protein production[13,342].

The nisin-inducible controlled gene expression (NICE) system is probably the most widely used inducible system for expression in *L. lactis*. Nisin is an antimicrobial peptide produced in some strains of *L. lactis*. The two-component system NisRK is triggered by nisin and consequently acts as transcriptional activator of the *nisA* and *nisF* promoter[343]. In the NICE system constructed by de Ruyter *et al*. in 1996, the *nisA* promoter is used to drive expression of the target gene while the genes *nisR* and *nisK* need to be integrated into the genome of the host strain[344,345]. Induction can be achieved with nisin or its analogs, and even with culture supernatant of nisin producing *Lactococcus* strains. This system is tightly controlled and provides relatively high protein expression levels[343,346]. The NICE system was used in the *L. lactis* study in Paper 3.

Table 3: Promoter systems used in *Lactococcus lactis*

Type	Promoter	Description	Ref.
Constitutive promoter	P_{32}	medium	347
	P_{21}, P_{45}	strong	348
	P_{8}	very strong	342
NICE expression system	P_{nisA}	nisin inducible, titratable, can only be used in strains with *nisKR*	344,345
Induction under environmental changes and stress conditions	P_{170}	auto-induction during transition to stationary phase, pH shift	349
	P_{Rro12}	temperature shift, works optimal at 42 °C	350
	P_{15A10}	phage attack	351
	P_{pstF}	phosphate starvation	352
Sugar-inducible promoters	P_{xylT}	xylose inducible, high expression levels	353
	P_{lac}	lactose inducible	354
Zinc-dependent promoters	P_{Zn}	suppressed by zinc, triggered by EDTA or zinc starvation	355
	P_{czcD}	activated in the presence of zinc, titratable	356

3.1.5 Gene expression systems for *Bacillus subtilis*

Being one of the most frequently used Gram-positive bacteria for protein production in biotechnology[357], the molecular toolbox for *Bacillus subtilis* and thereupon its expression systems have grown over the last years. The available promoter systems are summarized in Table 4.

Originally many constitutive promoters encoded on the genome of *B. subtilis* were used for recombinant protein production. Since the genome sequence of several *B. subtilis* species is available, promoter sequences can easily be identified[134,358] and can, after testing, be categorized in strong[106,359–361] and weak constitutive promoters[362]. Synthetic constitutive promoter, such as the Anderson promoter collection, have been proven to work in *Bacillus subtilis*[361]. Furthermore, promoter library approaches helped to extend the spectrum of promoter strength[106]. In this thesis protein production in *B. subtilis* was studied using the constitutive promoters P_{J23100}, P_{liaG}, P_{lepA}, P_{veg} and P_{32}, the latter of which is derived from *L. lactis* (Paper 3).

Inducible expressions systems with application in *B. subtilis* do exist. The SURE expression system (SUbtilin Regulated gene Expression) is regulated by subtilin, which exhibits homology to the well-characterised nisin used in the NICE expression system (see 3.1.4). Similarly, the lantibiotic triggers a two-component system, SpaKR, in *B. subtilis*[363]. Several strains containing the SpaKR system have been constructed for the SURE system[363,364]. However, the system is not as frequently used as the NICE system, because its inducer subtilin is not commercially available[364].

Similar to *L. lactis*, expression systems that respond to environmental changes have been developed for *B. subtilis*[365–368]. A cold-shock inducible system is especially interesting for protein production as expression at 25 °C has shown to be beneficial for protein production[369]. High expression levels can be achieved using a xylose inducible expression system. Xylose inducible expression systems are most frequently used for recombinant protein production in *B. subtilis* [370–372]. Other sugar-inducible systems exist. Two T7 expression systems have been rebuild in *B. subtilis*, one is IPTG inducible[373] whereas the second one responds to xylose[374].

Table 4: Promoter systems used in *Bacillus subtilis*

Type	Promoter	Description	Ref.
Constitutive promoter	P_{J23101}	very weak	361
	P_{liaG}	weak	362
	P_{serA}, P_{lepA}	strong	106,360
	P_{veg}	very strong	359
SURE expression system	P_{spaS}	subtilin inducible, titratable, can only be used in strains with *spaKR*	363
	$P_{spaSmut}$	subtilin inducible, can only be used in strains with *spaKR*, less titratable, higher expression compared to P_{spaS}	363
LIKE expression system	$P_{liaI(opt)}$	bacitracin inducible, titratable	375
Inducible expression system	$P_{xyl-tet}$	tetracycline inducible, very tight, medium expression level	376
	P_{liaI}	bacitracin inducible, titratable, nisin and vancomycin can also be used for induction	361,377
Induction under environmental changes and stress conditions	P_2, P_7	temperature shift, works optimal at 45 °C	368
	P_{des}	cold-shock inducible, works optimal at 25 °C	369
	P_{gsiB}	induced under glucose starvation, heat, low pH	365
	P_{APaseI}	repressed by high phosphate concentration	366
	P_{23S}	dependent on osmolarity, inducible with 2.5 % NaCl	367
Sugar-inducible promoters	P_{xylA}	xylose inducible, high expression levels, sensitive to carbon catabolite repression	370
	P_{sacB}	sucrose inducible, sensitive to carbon catabolite repression	378
	P_{mtlA}	mannitol inducible, strongly regulated by carbon catabolite repression	379
	P_{glvA}	maltose inducible, high expression levels, strongly regulated by carbon catabolite repression, a mutant version that is less susceptible to carbon catabolite repression exists	380,381
	P_{T7}	T7 RNAP under control of PxylA, xylose inducible	374
IPTG-inducible promoters	P_{spac}	IPTG inducible, co-expression of *lacI* for tight regualtion	382
	P_{grac}	IPTG inducible, higher expression levels compared to P_{spac}	372
	$P_{N25/O}$	IPTG inducible, co-expression of *lacI* for tight regualtion	383
	P_{T7}	T7 RNAP under control of P_{spac}, titratable up to 0.5 mM IPTG	373

3.2 Genome integration and genomic expression

With the advent of precise and efficient genome integration technologies, genomic-based protein production has gained attention. Compared to vector-based expression systems, expression of genome-integrated proteins reduces metabolic burden and provides genetic stability. In large-scale fermentation set-ups, genomic expression is more suitable and cheaper, as no antibiotics need to be supplemented to the culture medium[63]. Figure 14 illustrates the most commonly used genome

integration technologies for recombinant protein production in *E. coli*. Here, a comprehensive overview of these technologies is given. More detailed reviews about genome engineering technologies can be found elsewhere[384].

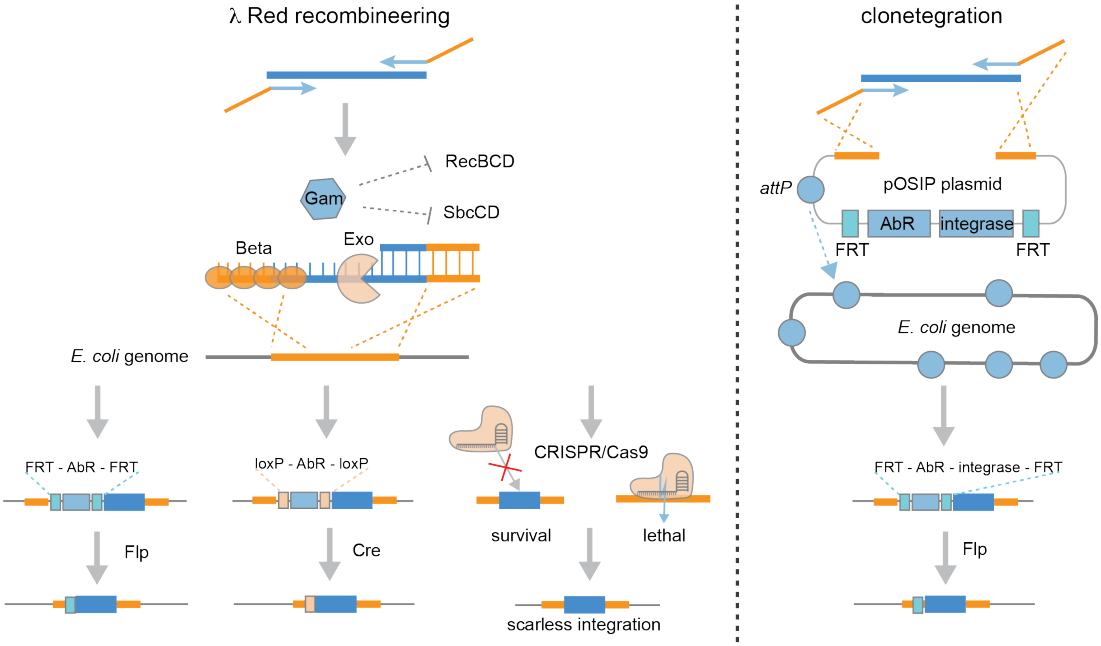


Figure 14: Genome integration techniques for chromosomal protein production. λ-Red recombineering is based on homologous recombination. The phage proteins Beta, Gam and Exo mediate and facilitate the integration process. Successful integrants can be selected by antibiotic resistance or by CIRSPR/Cas9 counter selection. The antibiotic resistance cassette can be removed via FRT/Flp or Cre/loxP-mediated excision. Clonetegration is a one-step cloning and integration approach. Integration can happen in six phage integration sites located on the *E. coli* genome. Successful integrants are selected on antibiotics. The antibiotic resistance marker and the integrase are excised by Flp.

λ-Red recombineering is a homologous recombination-mediated technique for genetic engineering used in prokaryotes. Recombineering can be carried out using either double-stranded DNA (dsDNA) or single-stranded DNA (ssDNA) templates. The system relies on three components, the Exo, Beta and Gam protein. Exo is a dsDNA-dependent endonuclease, which produces a single-stranded product for recombination. The single-stranded product is then protected from degradation by the Beta protein[385]. Thus, only Beta is needed for recombineering with single-stranded products[54]. The Gam protein inhibits the endogenous nucleases RecBCD and SbcCD, which degrade linear DNA fragments[55] (Figure 14). A publication by Mosberg *et al*. nicely describes the mechanisms behind λ-Red recombineering[385]. Generally, an antibiotic resistance gene is inserted simultaneously, to simplify the identification of recombinant clones after recombination. Selection on the respective antibiotic will reveal successful integrants. However, the selection marker needs to be removed after selection. This can happen via the Flp/FRT system from

Saccharomyces cerevisisae[386] (Figure 14). Alternatively, the Cre/loxP system from the bacteriophage P1 can be used[387] (Figure 14). However, both systems leave a scar in the genome which can prove problematic if several rounds of recombineering are performed in one strain resulting in false recombination events. Recent advances in CRISPR/Cas9-coupled recombineering, allow for markerless integration into the prokaryotic genome[68].

A simultaneous cloning and integration approach (clonetegration) was developed for integration into six phage sites located on the *E. coli* chromosome[388] (Figure 14). Clonetegration has been proven to be a fast and efficient integration method. However, antibiotic-based selection and subsequent Flp-mediated excision of an antibiotic resistance marker is required as well.

The antibiotic selection system developed in this thesis (Paper 1), can be employed to optimize protein production and to select for highly expressed gene variants on the genome of *E. coli* (see Appendix 2) using either λ-Red recombineering or clonetegration. The first one provides the means to select for optimized endogenous proteins.

3.3 Factors affecting recombinant protein production

E. coli is the most preferred and extensively studied system for the production of recombinant proteins in bacteria. Nevertheless, there are certain limitations for protein production that make expression of complex proteins a major challenge. For example, an average *E. coli* protein comes at a size of 300 amino acids, corresponding to about 33 kDa[389]. Proteins above 100 kDa can often not be expressed properly[390] and are secreted poorly[391] (Figure 15). Similarly, small proteins are often subject to proteolytic degradation[392]. In addition, many eukaryotic proteins frequently encompass codons that are rarely used in *E. coli* and thus slow down translation processes[393] (Figure 15).

Besides size and codon usage, toxicity presents a major problem. Many recombinant proteins need to be secreted to the periplasm for post-translational modifications but secretory and membrane proteins then often become toxic to the host cell (Figure 15). First of all, translocation of high levels of recombinant proteins hinders the translocation of endogenous proteins[394–397]. As a consequence of this, the respiratory capacity is reduced, which leads to a decrease in oxygen consumption rates and activates the ArcAB two-component system that mediates the transition from respiratory to fermentative metabolism[395,398]. Additionally, the accumulation of proteins in the cell's membrane

system has a toxic effect. However, the major problem presents the saturation of the membrane translocation machinery, which, firstly, causes aggregation of the proteins in the cytoplasm and their subsequent degradation. Proteins aggregate when hydrophobic side chains get exposed to the aqueous environment and interact with other hydrophobic residues on neighbouring chains to form non-functional aggregates. Logically, proteins with highly hydrophobic domains, such as membrane proteins, are more prone to aggregate. Aggregation is accompanied by changes in the protein structure therefore resulting in inactive proteins[399] (Figure 15). Large amounts of aggregates are sequestered as inclusion bodies. The latter are not necessarily unwanted, as they are usually highly homogeneous and can easily be separated from other cellular components[400]. Several protocols for inclusion body purification have been developed[401,402]. The aggregated proteins can be renaturated and solubilized to restore their activity[403,404]. However, protein extraction from inclusion bodies is a time consuming and tedious process, and soluble protein production is preferred.

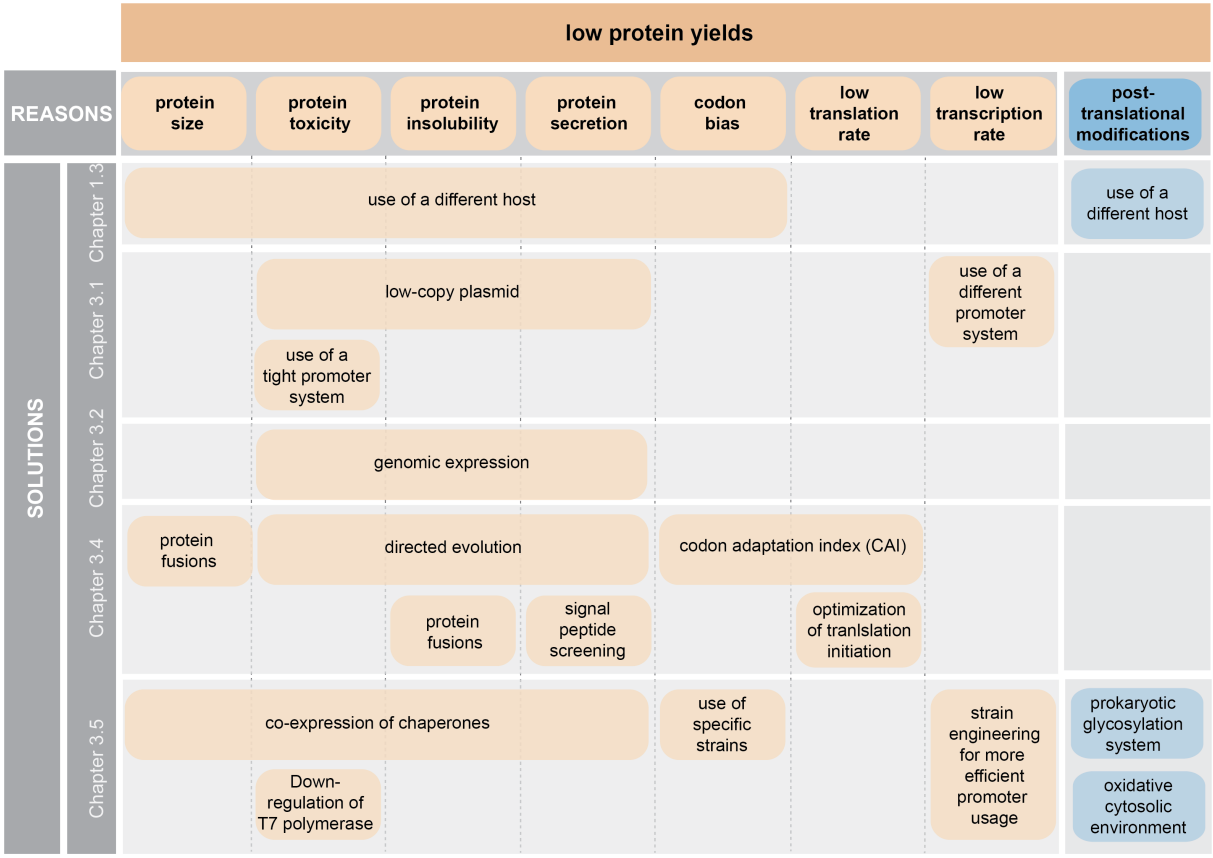


Figure 15: **Reason for and solutions to low protein yields in recombinant protein production.** Reasons for low protein yields are listed. Below possible solutions are provided. The grey bar on the left indicates in which chapter detailed information about particular solutions can be found.

3.4 Approaches for protein production optimization

Proteins can be engineered and modified to overcome the bottlenecks presented in section 3.3. The approaches presented below are connected with the protein production optimization tools that are developed in this thesis.

3.4.1 Protein engineering

Directed evolution

Modifying the nucleotide and amino acid sequence of a target gene has proven to be a successful approach to improve protein production. But in many cases the lack of pre-existing knowledge makes it impossible to predict which nucleotide and amino acid modifications in a target protein might have a positive effect on protein production[220]. Hence, optimized protein variants are often identified from sequence libraries. This approach is called directed evolution[405]. The generation of large sequence libraries can be achieved using error-prone PCR, gene shuffling or chemical mutagenesis[405–407]. The key goal with these approaches is to create great sequence diversity[406]. However, the application of directed evolution is not only limited to protein production optimization, including solubility and folding, it is also a powerful method for evolving single proteins with altered substrate specificity, enhanced activity and optimized functionality. Here, gene shuffling strategies have driven the success of directed evolution in the last decades resulting in hybrid enzymes, enzymes with altered or broadened substrate specificity and pH stable protein variants, just to name a few[408–414]. More targeted approaches can be employed by using algorithms that design shuffled libraries from non-homologous genes *in silico,* e.g. the SCHEMA or SCOPE algorithm[415–418].

A screening and selection method is required to efficiently identify the optimized mutant from these large libraries. Commonly, fusions to the green fluorescent protein (GFP) are used[419]. In this thesis, we developed a selection system with a huge screening capacity due to a growth-based assay that could be implemented as a screening tool for directed evolution of proteins for production optimization (Paper 1 and 2).

Codon adaptation index (CAI)

It is also possible to engineer the DNA coding sequence of a protein without actually changing the amino acid sequence. There is a repertoire of codons for each amino acid and certain codons are typically more frequently used than others. The codon adaptation index (CAI) is the most

widespread technique to score the abundance of a codon but other techniques exist[420]. CAI calculates the statistical codon usage relative to the codon usage in highly expressed genes to predict protein expression levels[421]. Hence, proteins that use abundant codons score highly in the CAI whereas those that use rare codons will score low in the CAI. The CAI can affect translation elongation. Introducing more abundant synonymous codons can increase production by up to 1000-fold[422] but can also alter protein folding and therefore activity[421,423,424]. Questioning this common understanding of codon bias, Kudla *et al*. engineered a synthetic library of 154 *gfp* genes that varied randomly at synonymous sites. When expressed, protein production varied by 250-fold but no correlation between CAI and expression level was found. The effect of CAI seems to play a role in global fitness rather than in production optimization. Those findings were confirmed in later studies[425–427]. Goodman *et al*. even found that rare codons at the 5′ have a beneficial effect on expression[426]. More recent studies, using ribosome profiling techniques, uncovered some relationship between internal Shine-Dalgarno sequences and ribosome pausing. For this reason, the exclusion of SD-like codons can be beneficial for protein production[428]. Gene libraries that vary at synonymous sites can be screened with the here developed antibiotic selection system for their expression performance (Paper 1).

3.4.2 Protein fusions

Fusing the target protein to another soluble protein can increase folding and solubility[390]. This way, one can protect small proteins from proteolysis and large proteins from aggregation[429–431]. Fusion proteins or fusion tags can also facilitate protein purification[432], and some fusion proteins can be used as expression reporters. Nevertheless, fusion proteins must be chosen and engineered carefully. For instance, secretory and membrane proteins contain information in the N-termini that is required for correct targeting. Although there are trends that some fusion partners are on average performing better than others, there is no guarantee that a given tag will work with a protein of interest.

Green fluorescent protein (GFP)

The green fluorescent protein (GFP) is commonly used in molecular biology as it is a stable protein that can easily be monitored in whole cells. GFP is used as fusion reporter for monitoring expression levels, protein folding, localization and membrane protein topology. Further, fusions to GFP were found to improve the solubility of a protein and protein purification strategies[433,434].

Waldo and co-workers developed a folding reporter-vector in which a target protein is expressed with GFP as a C-terminal fusion. Conveniently, it was shown that the GFP reporter does not fold properly if the target protein aggregates. As a result GFP does not fluoresce. The reporter can be used to assay, for example, the solubility after directed evolution experiments[435]. In addition, the GFP folding reporter does not fluoresce upon secretion into the periplasm, and can thus be used as a reporter for protein localization[436]. Building on these findings, Drew *et al*. employed the GFP folding reporter for mapping the topology of membrane proteins[437] and to monitor membrane protein expression levels[438]. As a continuation of the platform developed by Drew and colleagues, the GFP folding reporter was applied to 601 inner membrane proteins of *E. coli* and their membrane topology was analyzed[439]. Split-GFP variants present good alternatives for topology mapping and protein tagging in general[440]. Commonly the protein is split between the tenth and eleventh β-strand. The resulting GFP_{1-10} fragment remains non-fluorescent until complementation. The resulting GFP_{11} fragment constitutes a 16 amino acid short peptide, which can easily be fused to a target protein. Besides split-GFP, different variants of GFP with enhanced properties were developed. Superfolder GFP, which is a very stable version that also folds properly in the periplasm presents here the most prominent example[441].

The implementation of fluorescent proteins has also enabled the use of fluorescence-detection size-exclusion chromatography to evaluate the quality of purified proteins, e.g. as a result of different detergents used for solubilisation, and to track higher order structures in purification processes[442,443]. Moreover, the combination of GFP-fusions and fluorescence activated cell sorting (FACS) enables the selection of high producing variants from expression libraries. Skretas and Georgiou screened an *E. coli* transposon library with the latter approach and identified a variant with substantially increased production of the membrane-integrated human central cannabinoid receptor (CB1) [444].

In this thesis, we made use of the various benefits of membrane protein-GFP fusions. Two membrane proteins of the GFP platform[445] were used to test and characterize the tunable antibiotic selection system TARSyn, developed in Paper 1. Additionally, a protein from the GFP platform was used to characterize the impact of all SEVA-Linker plasmid backbones on protein production, shown in Paper 4.

β-lactamase

The ampicillin resistance-conferring protein β-lactamase can also be employed as a protein fusion. Although a fusion with β-lactamase may not improve protein production as such, it is a useful screening tool for optimized production. Skretas and Georgiou used membrane proteins fusions to β-lactamase to find enhanced production of membrane proteins when screening the ASKA library, a plasmid library encoding all known *E. coli* open reading frames[446]. As β-lactamase only acts in the periplasm, screening for optimized secretion is another possible application. The laboratory of DeLisa recently showed that β-lactamase can be used to screen for optimized extracellular secretion through strain engineering[447]. This thesis exploits extensively the use of β-lactamase as a reporter for protein translation in the antibiotic selection system. However, we did not use β-lactamase as a direct fusion partner, instead we translationally coupled the gene of interest with the reporter.

Other examples of fusion partners, which mostly support solubilization and folding of the target protein, are listed in Table 5.

Table 5: Fusion proteins for recombinant protein production in *E. coli* and their characteristics

Fusion protein	Characteristics	Ref.
Maltose binding protein (MBP)	• very big (42 kDa) addition to the target protein • translocation of the target protein into periplasm (due to *malE*-SP) • increases solubility of target protein • has intrinsic chaperone activity, reduces toxicity of target protein • prevents degradation of target protein • serves as affinity tag for purification *via* amylose	448–457
Glutathione S-transferase (GST)	• smaller (26 kDa) addition to the target protein • rather poor solubility enhancer • serves as affinity tag for purification *via* glutathione	458–461
Ubiquitin and small ubiquitin-like modifier (SUMO)	• small addition to the target protein • increases solubility of target protein • increases stability of target protein • additional Histidine-tag needed for purification	461–465
N-utilizing substance A (NusA)	• very big (55 kDa) addition to the target protein • increases solubility of target protein (best soluble protein of *E. coli*) • "attracts" chaperones for proper folding • slows down translation for proper folding	466–469
Thioredoxin A (TrxA)	• increases solubility of target protein • good fusion partner for crystallographic studies, as it tends to form crystals itself	460,461,467, 470,471

Affinity tags for purification

Smaller peptide tags are probably less likely to interfere with protein structure, activity and characteristics and therefore do not necessarily need to be removed. On the other hand, they mostly serve for purification purposes and likely do not support solubilization. They are listed here, as the choice of affinity tag can greatly influence the protein yield after purification. An optimal production scenario includes optimal purification conditions for the process to be economically viable. The most frequently used affinity tags are listed below (Table 6), but many other fusion partners, such as the calmodulin-binding protein, the chitin-binding domain or the cellulose-binding domain exist and were nicely reviewed by Terpe[432]. In the studies presented in this thesis we employed fusions to His-tag and Strep-tag for successful protein detection and purification.

Table 6: Commonly used affinity tags for protein detection and purification

Affinity tag	Characteristics	Ref.
poly-histidine tag (6xHis-/ 8xHis-tag)	• Affinity purification through Ni2+-NTA resin or immobilized metal-affinity chromatography (IMAC)	459,472,473
Strep-tag	• Affinity purification through binding to strepavidin and elution with biotin	474–479
FLAG-tag	• artificial designed peptide, very hydrophilic • through internal cleavage site elution under non-denaturating conditions possible • affinity purification through FLAG-antibody binding • affinity pull-down	480–483
HA-tag	• used for coimmunoprecipitation (co-IP) • especially strong epitope for purification and detection	484
cMyc-tag	• used for coimmunoprecipitation (co-IP) • used for immunofuorescence assays • can interfere with protein translocation	459,485

Despite all their advantages fusion proteins can compromise the expression, solubility and bioactivity of a protein and should be only sensible when the protein fusion itself constitutes the desired product. In industrial applications fusion proteins are usually avoided due to additional costs and time required for removal of the fusion partner. Further, for pharmaceutical applications additional sequences are unwanted due to safety regulations[486]. Tags can be removed by e.g. protease cleavage, but most proteases leave a few residues behind. Last but not least, one risks with the cleavage of a fusion partner the loss of solubility of the target protein[487].

3.4.3 Optimization of protein secretion

Secretion of recombinant proteins to the periplasmic space of *E. coli* has several advantages over intracellular production. This includes a higher solubility, stability and enhanced activity, if disulfide bonds need to be formed, and easier downstream processing, as proteins can simply be obtained by osmotic shock or permeabilization of the cell wall[253,261,263,488,489]. However, the choice of signal sequence (ss) impacts greatly the efficiency of secretion and decides via which route a proteins gets translocated (see Chapter 2). In the model organism *E. coli*, the route for many endogenous periplasmic proteins is known, which simplifies the selection of a signal peptide[242,243]. Often post-translational translocation is preferred as the risk for aggregation and misfolding is lower[130,490]. However, for other applications, e.g. phage display, co-translational translocation has proven to be most efficient[242]. Commonly used signal sequence (ss) are PelBss, MalEss, PhoAss and signal sequences of the Omp proteins that mediate post-translational translocation, whereas DsbAss, TolBss and TorTss drive co-translational translocation[244]. Proteins utilizing the TorAss will be translocated via the Tat pathway[130] (Figure 16). Direct protein fusions to a leaderless β-lactamase gene can be employed to identify optimal signal sequences in a growth-based selection assay, similar to the selection assay that was developed in this thesis (Paper 2). Additionally, the TIR optimization approach that is employed in Paper 1-3 will, when applied to signal sequences, result in universal optimized scaffolds for periplasmic protein production.

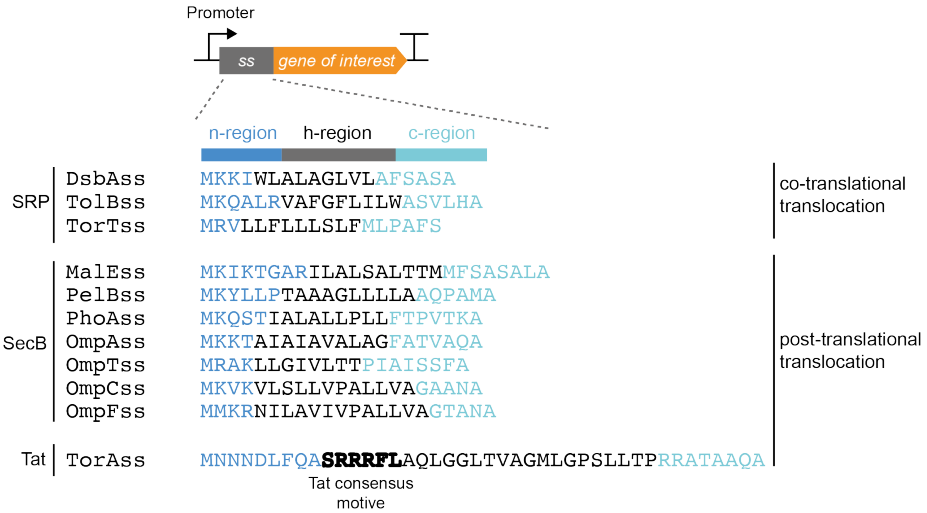


Figure 16: Coding sequences for commonly used signal sequences (ss). Signal sequences contain three regions, the n-region (blue), the h-region (black) and the c-region (light blue), and are fused to the N-terminal of a gene of interest (orange). Signal sequences encode whether a protein will be translocated co- or post-translationally via SRP-dependent, SecB-dependent or the Tat pathway. (Figure adapted from Steiner *et al.*[242] and Choi *et al.*[130])

3.4.4 Optimization of translation initiation

It has been shown that the rate-controlling step in protein synthesis is translation initiation [129,491,492]. The nucleotide sequence central for this process is the so-called translation initiation region (TIR) that comprises the Shine-Dalgarno (SD) sequence, a linker region between the SD sequence and the translational start site and approximately the first five codons of the gene coding sequence and optional translational enhancers such as A/U-rich sequences[189,493,494]. Details about the mechanisms behind translation initiation can be found in Chapter 2.2.1. The efficiency of translation initiation mostly depends on four factors (Figure 17): The nature of the SD sequence, the start codon for translation, the length and sequence of the linker region between SD sequence and the start codon, and the free energy (ΔG) of folding of the mRNA corresponding to its secondary structure around the translation start site.

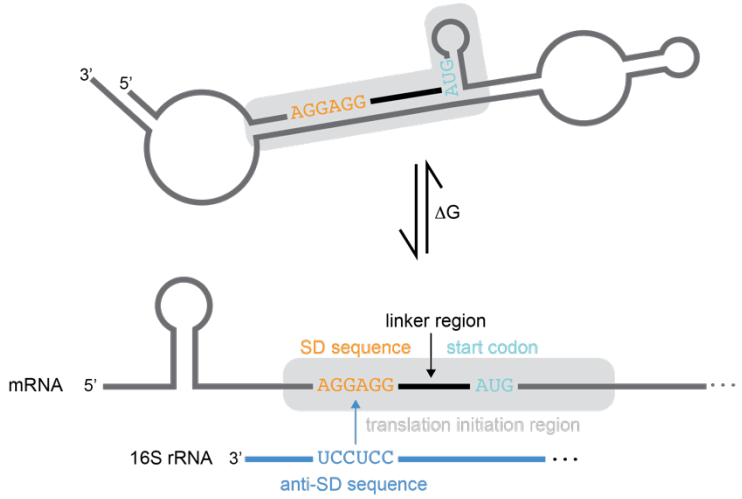


Figure 17: The translation initiation region. The translation initiation region (light grey) spans from the SD sequence to the 5th or 6th codon of the gene. The efficiency of translation initiation depends on the Shine-Dalgarno sequence (SD sequence, orange), which needs to be recognized by the 16S rRNA, the linker region between SD and start codon (black), the start codon (turquoise) and the secondary structure of the mRNA which needs to be resolved (ΔG). (adapted from Reeve *et al.*[493]).

All four points have extensively been studied in the last decade and many approaches have been developed that optimize translation initiation and therefore protein production levels. One of the most important steps in translation initiation is the binding of the 30S ribosomal subunit to the mRNA, where the 16S rRNA pairs with the SD sequence of the mRNA[185,495]. Nucleotide changes in the SD can therefore affect protein production levels by as much as 600-fold[496] and up to 100,000-fold (personal communication). Vimberg *et al.* tested the effect of SD variation on the level of protein synthesis. They found that at 37 °C the consensus SD sequence AGGAGG performed best in *E. coli*[189]. Most commercially available expression vectors contain this consensus SD sequence, hence it is difficult to improve translation by changing it. Nevertheless, for metabolic engineering

applications a reduction in expression levels is sometimes favoured. To predict the performance of weaker SD sequences, one can make use of the prediction software EMOPEC[497]. However, as translation efficiency does not only depend on the SD sequence it is hard to gauge whether a construct with altered SD will perform as predicted. To address this problem, Mutalik and co-wokers developed a standardized bicistronic design (BCD), comprising a small peptide that can easily be incorporated in front of any gene of interest. The BCD contains two SD motifs, one driving translation of the small peptide and a second SD, encoded in the small peptide sequence, driving translation of the target gene. Variation was introduced in the second SD and selected BCD sequences showed a 600-fold variation in expression[496]. Using such BCD elements makes translation initiation rates reliable and transferable between target genes. Our own experience taught us that expression levels of already optimized constructs could still be doubled by utilizing the best-performing BCD sequence (unpublished data). Marino *et al*. tried a similar approach by designing transcriptional fusions of target gene and the proteins Mistic and YebL. Expression was enhanced but not comparable to the biscistronic design of Mutalik and co-workers[498].

The second factor influencing translation initiation is the initiator codon. Most commonly the start codon is AUG but sometimes GUG and UUG and rather rarely CUG or AUU are used[33,499–502]. The use of AUG is most efficient for translation initiation as all three nucleotides base pair with the initiator fMet-tRNA[503]. A recent study by Hecht and colleagues showed that in theory all 64 codon can serve as an initiator codon in translation initiation, as long as a strong SD sequence is used[188]. A study in this thesis makes use of the non-canonical AUU start codon to down-regulate expression of a translationally-coupled reporter gene, which can be useful if a gene of interest is already highly expressed and the detection range of the reporter is already saturated (See Paper 1).

The region between the SD sequence and the AUG start codon was long thought not to play an important role in translation initiation. However, Egbert and Klavins showed that by varying the length of this region, gene expression levels varied over a 1,000-fold range[504]. The optimal spacer length was shown to be 5 nucleotides[505] but can vary with the context of the SD sequence[494]. When expressing a gene of interest, the construct is often assembled on an expression vector using various cloning techniques. Independent from which cloning technique is used, an arbitrary junction will be created between the vector backbone and the coding sequence, exactly the sequence between the SD sequence and the start codon, the context of which will influence expression levels[107,506]. Modifications applied to this region might also change the secondary

structure of the mRNA. An algorithm to predict mRNA folding around translation start site has been developed[507], however, it was shown that expression variability cannot completely be explained by the free energy associated with mRNA folding and seems to be context specific in a way that is not completely computationally predictable[107].

The creation of sequences libraries, can help to find the optimal TIR sequence for high-level expression of a gene directly under the conditions in which it will be expressed. Sampling from TIR libraries yields clones with very large differences in expression. However, the specific randomized region decides on whether experimental evaluation is possible (Figure 18). In so-called RBS libraries, variation is introduced in the SD sequence and its direct surroundings. The resulting library includes a high frequency of poorly performing constructs, due to suboptimal SD sequences[508] (Figure 18). When only manipulating the region downstream of the SD sequence and sampling all synonymous combinations for the second and third codon, the library size is reduced and experimental evaluation possible (Figure 18).

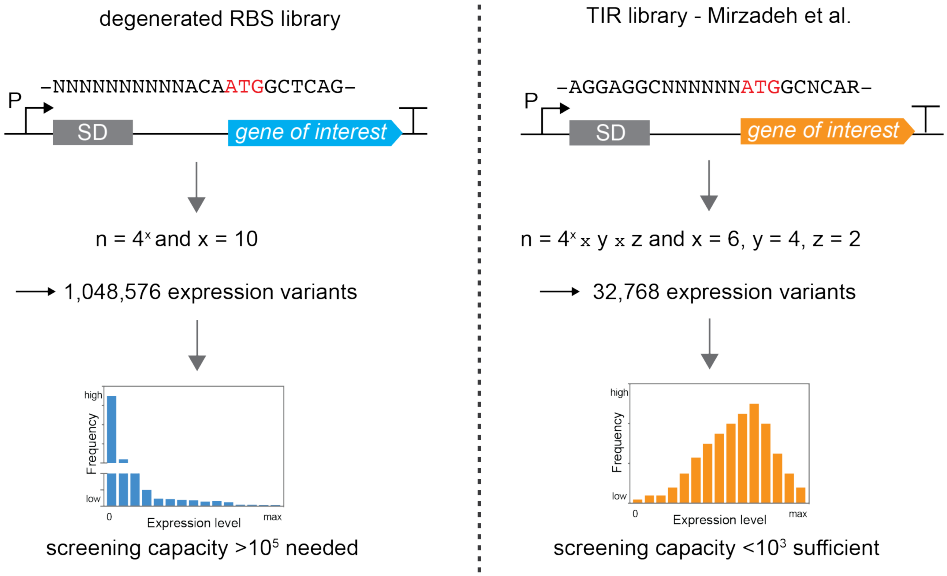


Figure 18: Different DNA sequence libraries for the optimization of translation initiation. So-called RBS libraries generally introduce variation in the SD sequence and its surroundings, creating numerous variants with low expression levels (left). If 10 nucleotides (x=10) are randomized completely, 1,048,576 expression variants are created. Screening capacity needs to be high to be able to identify the good candidates. The presented data is imaginary but resembles observations by Jeschek *et al.*[508]. When only a part of the translation initiation region (TIR) but the SD sequence is kept optimal, the library size and the frequency of bad solutions is reduced (right). Mirzadeh *et al.* introduce variation in the 6 nucleotides (x=6) upstream of the start codon. The second and the third codon is changed to all possible combinations (y=4, z=2). The presented data is imaginary but resembles observations by Mirzadeh *et al.* [107].

The latter approach was described by Mirzadeh *et al*. The authors found that protein production levels were affected by up to 1000-fold[107]. In this thesis TIR sequence libraries as described by Mirzadeh *et al*. have been utilized to optimize protein production. These libraries were constructed to apply TARSyn, a synthetic biology approach for selection of highly expressed gene variants developed and applied in Paper 1 to 3. However, due to its high screening capacity, we are confident that TARSyn can be used to screen degenerated RBS libraries.

Supporting the previous example, other studies found synonymous codon substitutions in the second and third codons of the target gene to have a large effect on protein production levels[509–512]. However, it is not clear whether those codon substitutions affect mRNA folding around the translational start site[426,513,514] and therefore translation initiation, or the ribosomal speed and would therefore affect translation elongation[515,516]. Most likely, both processes are affected. Cheong and co-workers enhanced expression levels by creating an expression library where the first ten codons of a gene of interest were substituted by all possible combinations of synonymous codons. This approach resulted in up to 530-fold increase in expression[517]. Nørholm at al. concluded that the 5'-coding region is the major region of the protein coding sequence that influences translation efficiency. Their conclusion is based on a study of two difficult-to-express membrane proteins. Codon optimized sequences of both genes were compared to variants where the codons close to the AUG start codon were substituted with synonymous codons. The latter showed a substantial improvement in production[512].

Very importantly, this thesis provides a new comprehensive study that shows that such library approaches for translation initiation optimization can also be applied in the Gram-positive hosts *B. subtilis* and *L. lactis* (Paper 3).

Three different computational tools for the prediction of the translation rate have been developed. All of them use biophysical models to calculate and design optimal translation initiation regions. The RBS Calculator, launched in 2009, was the first of its kind[108]. New results and insight in the translation initiation mechanism led to an on-going development of the RBS Calculator[518,519]. The latest version, RBS Calculator 2.1, was released in September 2017[520]. In addition to predicting translation rates, the RBS Calculator offers the means to design RBS libraries, which is mostly applied for pathway optimization and less for protein production optimization[521–523]. The RBS Designer and UTR Designer are similar computational tools[524,525]. However, as this thesis focuses

chiefly on experimental approaches for protein production optimization, the reader is referred to the thorough review by Reeve *et al*.[493] for further details on the subject matter.

3.5 Optimization of the protein production host

3.5.1 Codon usage

In the previous section codon abundance and the use of CAI was discussed. As an alternate solution to CAI, tRNAs for rare codons can be co-expressed to circumvent the codon bias problem. Usually rare codons correlate with rare tRNAs[526]. The pR1952 and pSJS1240 plasmids[527,528] or the pRIG plasmid[529], for example, encode two and three rare *E. coli* tRNA genes, respectively. Those plasmids have proven to be useful in the expression of AT-rich genes. Commercially available tRNA plasmids include pRARE and pLysSRARE2 from Novagen which can be purchased in the Rosetta, and Rosetta2 strain (Novagen) and carry tRNA genes for seven rare codons. The BL21-CodonPlus strain from Stratagene also supplements rare tRNAs and comes in different variants, such as CodonPlus-PL and CodonPlus-RIL that can be used to overcome biases of GC- and AT-rich genes, respectively[390]. Both commercial expression platforms have been used successfully, for example in the Human Protein Atlas project[530]. One should, however, be aware, that the presence of additional tRNAs can also have a negative effect on recombinant protein production, as reported by Søgaard and Nørholm[531].

3.5.2 Co-expression of molecular chaperones

The over-expression of molecular chaperones is another strategy that has been used to improve the yield in the production of recombinant proteins. Chaperone functions are described in more detail in Chapter 2. Host chaperone accumulation can be induced by heat, osmolytes or membrane fluidizers and does not necessarily require plasmid-mediated over-expression[488,532]. Recent experiments indicate that the optimization and selection approach that is applied in this thesis, can also be used to optimize translation initiation of the chaperones on the genome of *E. coli*, which results in elevated levels when production is induced.

Over-expression of the cytoplasmic chaperones DnaKJ-GrpE, GroEL/GroES and TF has been used in several studies to enhance yield and solubility of several human proteins, including the very difficult-to-express class of GPCRs[533–535]. Opposing these previous studies, Skretas and Georgiou

found that the disruption of DnaJ substantially increased production of another GPCR[444]. Similarly, Nannenga and Baneyx found that the deletion of TF tripled production of two membrane proteins[536]. However, the involvement of DnaKJ in protein degradation, and competition of TF and SRP for binding to the emerging polypeptide chains, can provide reasonable explanations. Nevertheless, this demonstrates that co-expression of chaperones requires specific testing for each target protein. Unfortunately, no screening platform for such purpose exists. A few pACYCDuet-1 derivatives for the co-expression of chaperones are available *via* Addgene. Those plasmids originate from a study for soluble production of human lysozyme in *E. coli*[537]. At this point, it is important to mention that the antibiotic selection system developed in this thesis, monitors the translation rate and cannot distinguish between correctly folded and misfolded, aggreagated proteins. The implementation of a simultaneous folding reporter would be of high interest for the screening of chaperone co-expression.

As controversial as the results are for cytoplasmic and membrane proteins, they appear to be quite consistent for the optimization of periplasmic proteins. The co-production of FkpA, Skp and SurA can substantially improve production yields of secretory proteins, such as antibodies and scFv antibody fragments but also other protein classes[490,538–540]. Yields were improved up to 36-fold in those studies. The co-production of DsbA and DsbC when producing proteins, which are rich disulfide bonds, has led to success[541,542543]. For example, the production of an IgG antibody was enhanced 3-fold, when over-expressing DsbA or DsbC[544]. Schlapschy *et al*. constructed a helper plasmid that combines all previous efforts and encodes four established periplasmic chaperones, DsbA, DsbC, FkpA and SurA. Its positive effect was demonstrated using the human plasma retinol-binding protein (RBP) as well as the extracellular carbohydrate recognition domain (CRD) of the dendritic cell membrane receptor DC-SIGN[545]. Nonetheless, co-expression of chaperones can also have side effects, such as reduced yield due to reduce solubility or degradation, reduced specific activity or growth inhibition. Side effects of chaperone co-expression were nicely reviewed by Martinez-Alonso and co-workers[546].

In *B. subtilis* and *L. lactis* the co-production of chaperones showed a positive effect as well. The surface anchored chaperone-like protein PrsA from *B. subtilis* and the peptidyl-prolyl-cis/trans-isomerase PmpA from *L. lactis* both optimized production of secreted proteins when over-expressed[547–549].

3.5.3 Strain engineering for a more effective T7 expression system

The T7 polymerase can be a big bottleneck, especially when produced proteins or compounds tend to be toxic for the cell. To test the β-carotene pathway in the SEVA linker series, presented in Paper 4 of this thesis, we were in need of an optimized T7 system. The laboratory of Christopher Voigt reduced T7 RNAP toxicity by utilizing a weak Shine-Dalgarno sequence and a GTG start codon. Additionally, an N-terminal lon-mediated degradation tag was inserted. A spontaneous point mutation in the polymerase active site (R632S) occurred and showed a positive effect. The engineered T7 RNAP can be obtained on a plasmid[74]. For our study we successfully integrated the mutant T7 RNAP into the genome of *E.coli* and obtained β-carotene production of about 4 mg/mL.

Alternative strategies have been developed. Lemo21(DE3) is a derivative of BL21(DE3) that has the T7 lysozyme expressed under control of the rhamnose inducible P_{rhaBAD} promoter. Thus, the level of T7 lysozyme and thereby also the activity of the T7 RNAP can be carefully titrated. By varying inducer concentrations, optimal protein production yields can be achieved e.g. by reducing toxicity[550]. The construction of pLemo was inspired by the Walker strains. Evolved strains such as the Walker strains C41(DE3) and C43(DE3) can improve membrane protein production, similar to the recently published Mutant56(DE3)[550–552]. The first two BL21(DE3) derivatives contain, among other potential mutations, a mutation in the lacUV5 promoter that decreases the production of T7 RNAP and therefore leads to decreased transcription rate[550]. The latter one contains an additional deletion in FryA, a putative transporter subunit, which improved strain fitness.

To obtain a similar effect, Kim *et al*. constructed a titratable T7 system by putting a mutant $LacI^{V192F}$ repressor under control of a rhamnose inducible P_{rhaBAD} promoter. The mutant LacI variant binds to the lac operator site but does not bind to inducer molecules like IPTG. When $LacI^{V192F}$ binds to the operator site, transcription will be blocked. Through this, the activity of T7 RNAP can be carefully titrated. It should be noted that the use of the pLemo is limited to T7 RNAP-based expression systems whereas the LacI system is applicable to all other lacO-based expression systems that use endogenous *E. coli* RNAP[553]. The antibiotic selection system that is presented in this thesis provides the means to screen for optimal induction of the mutant LacI and of T7 lysozyme, in the case of pLemo, in liquid culture MIC assays.

3.5.4 Strain engineering for post-translational modifications

Production of a target protein can be optimized by utilizing the synthetic biology approaches developed in this thesis. However, sometimes additional post-translational modifications are required to obtain an active protein. In those cases, it is necessary to co-express the appropriate helper plasmids during the optimization process, or, use the appropriate strain in the optimization process (unpublished data). Therefore, a short insight into the options for post-translational modifications will be given.

Disulfide bond formation

If the reductive cytoplasmic environment is transformed into an oxidative environment, disulfide bonds can be formed in the cytoplasm and secretion is not necessary. Several *E. coli* strains have been constructed for this purpose. The thioredoxin and glutaredoxin pathway are involved in the reduction of disulfide bonds that form in the *E. coli* cytoplasm[554]. Its key enzymes, *trxB* and *gor*, respectively, have been knocked out or mutated in some strains to increase disulfide bond formation in the cytoplasm[555]. Such strains are commercially available, e.g. the Origami or Rosetta-Origami strains from Novagen. Additionally to the *trxB* and *gor* mutations, SHuffle strains, available from New England Biolabs, express a leaderless DsbC protein which thus resides and is active in the cytoplasm[556]. The SHuffle strain was successfully applied to the cytoplasmic production of a full-length antibody and various human proteins[557,558]. Very recently, the laboratory of DeLisa successfully transferred the whole periplasmic oxidation pathway into the cytoplasm by solubilizing the membrane protein DsbB and co-expressing a leaderless DsbA protein[559]. The system outperformed the SHuffle system in the production of a mouse urokinase (uPA) containing 12 disulfide bonds.

Hatahet, Nguyen and colleagues developed a plasmid-based system that allows for the formation of disulfide bonds in the cytoplasm without disruption of the reducing pathways. The plasmid expresses the yeast mitochondrial protein Erv1p and the mature disulfide isomerase DsbC from a pLysS backbone and was later named pCyDisCo (cytoplasmic disulfide bond formation in *E. coli*)[560]. In contrast to other disulfide bond catalysts, Erv1p is able to function independently of an intermediary protein (e.g. DsbA for the disulfide bond catalyst DsbB)[561–563]. Therefore, co-expression of Erv1p in the *E. coli* cytoplasm is sufficient to increase the activity of recombinant proteins by 1000-fold. Compared to the Δ*trxB* Δ*gor* origami strain the activity was doubled when over-expressing Erv1p[564,565].

Glycosylation

For a long time, glycosylation was thought to be a solely eukaryotic feature. The discovery of an N-linked glycosylation system in the gram negative bacterium *Campylobacter jejuni*, the *pgl* gene cluster, opened up the possibility to synthesize glycoproteins in prokaryoates[566]. Expression of the system's oligosaccharyl-transferase PglB in *E. coli* led to glycosylation of its substrate AcrA[567], and in a more applied study even to glycosylation of a scFv antibody fragment, which increased biophysical and pharmacokinetic properties[568]. In 2012, the laboratory of DeLisa published a major breakthrough in bacterial glycoprotein production. Expression of four yeast-derived enzymes, involved in the eukaryote's glycosylation process Alg1, Alg2, Alg13 and Alg14, and PglB in *E. coli* resulted in the *in vivo* formation of a $Man_3GlcNAc_2$ glycan, the core structure to all human N-glycans[569]. However, the subsequent composition of the produced glycan and the efficiency of glycan attachment to a protein remain major challenges for glycosylation in prokaryotes. Additionally, a difference in the consensus sequence recognized by PglB and the eukaryotic consensus sequence presents a bottleneck in the development of *E. coli* as a glycoprotein production host[570]. An excellent review on the details of protein glycosylation in bacteria was published in 2010 by Nothaft and Szymanski[571].

3.6 Screening for optimized protein production

Available screening techniques vary in their capacity, sensitivity and selectivity. The ideal screening assay should be (i) highly sensitive, which enables the detection of small differences, (ii) should cover a large dynamic range, which prevents saturation of the system, and (iii) should respond in a linear range of detection[572]. The correlation between an input and an output of a screening assay is defined in its transfer function (Figure 19).

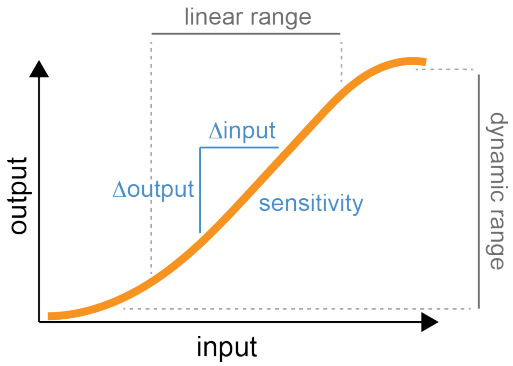


Figure 19: The transfer function. Correlation between the input, i.e. protein concentration, and the output response are important. The linear range, the dynamic range and the sensitivity (Δoutput/Δinput) define a reporter.

Construction of the reporter system for protein production

When optimizing production of a protein of interest, a reporter protein helps to identify the optimized candidates in screening assays. The choice of reporter protein decides on which screening techniques can be used later on (Table 7). A reporter is not always essential, but often enables high throughput. Commonly used reporters are fluorescent proteins, proteins that form or bind to pigments, or antibiotic resistance markers. The production of the reporter protein needs to be connected with the target protein production. In most cases, the target protein is fused directly to a reporter protein or a split-reporter protein as described previously (Figure 20A, B). An attractive alternative is the coupling of target and reporter gene on translational level. Translational coupling setups avoid the physical fusion of target and reporter protein. Translational coupling is a natural phenomenon occurring in prokaryotes, for example in the attenuation of the tryptophan operon[573]. Such setups can be constructed utilizing secondary structure forming RNA devices. Masking of the SD sequence of the reporter gene in an mRNA secondary structure, makes translation of the reporter gene dependent on the translation of the target gene. Only if the target gene is translated, the helicase activity of the ribosome will unwind the secondary structure and expose the SD sequence for ribosome binding (Figure 20C).

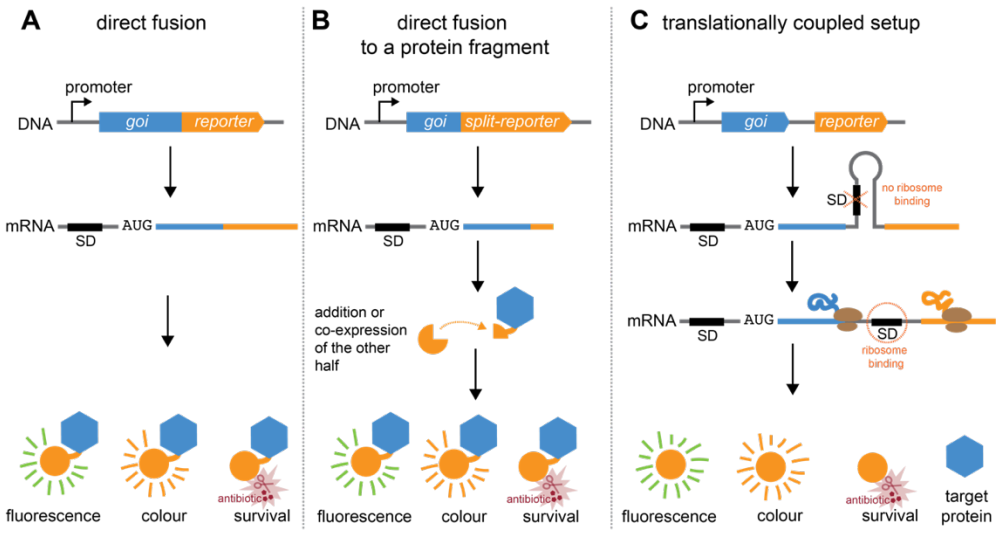


Figure 20: Direct and translationally coupled protein fusions. A gene of interest (*goi*, blue) can be fused directly to a reporter gene (orange) (A). Reporter proteins can be split into two parts, which upon reassembly exhibit the same phenotype as their non-split counterparts. The gene of interest can be fused to a split-part of a reporter protein. The complementary split-half needs to be supplemented or co-expressed (B). (C) Target gene and reporter gene can be translationally coupled. Translation of the reporter depends then on translation of the target gene.

Mendez-Perez *et al*. designed the first synthetic RNA hairpin structure[574]. Yet, the target gene needs to be C-terminally His-tagged to make use of this coupling device. In the first study of this thesis

(Paper 1) an alternative translational coupling device is developed, which no longer requires an artificial extension of the target gene. Translationally coupled fusions are especially useful if a growth-coupled phenotype, e.g. an antibiotic resistance marker should be utilized for screening. A previous study found that many antibiotic resistance marker form oligomeric complexes and as a result can drag the target protein in a misfolded state[575]. Independent of whether direct fusions or translationally coupled setups are employed, any reporter proteins with colorimetric or fluorescent phenotypes can be used for screenings.

Screening and selection

Which screening approach can be applied, depends on the phenotype of the reporter. Screening approaches are summed up in Table 7. Screening on LB agar plates or in microtiter plates can be very time consuming and laborious but is needed in case of a colourimetric phenotype. The throughput of manual screening is rather low with less that 10^5 clones per day. In the presence of a fluorescent phenotype, higher throughput can be achieved by using fluorescence-activated cell sorting (FACS) (up to 10^9 cells per day) or droplet micro-fluidics[88,576–580]. While being extremely efficient, FACS and droplet micro-fluidics are expensive platforms to purchase and maintain, and specific training is required to operate such platforms. The most powerful screening method in terms of throughput is to link protein production to fitness or survival of the cell[88]. This is an easy task if the target is essential for growth. If not, the production of the target protein can be linked to a growth-essential factor, such as an auxotrophic marker or an antibiotic resistance gene[572,581]. The drawback of these systems is that selection becomes difficult in strains where the desired protein is already produced to decent levels and has to be further improved[582]. If antibiotics need to be supplemented in g/L, screening is not economically feasible anymore. This thesis addresses the drawback by outlining the design of novel translational-coupling devices, which tune the coupling efficiency, thereby extending the efficacy of the used reporter gene.

Estimating gene expression levels by screening small molecule products

In the case that the target protein performs an enzymatic reaction and the product can easily be measured, the product formation itself can be used to indicate expression levels. Inconspicuous small molecules can be converted into a detectable output through transcriptional activation of a conspicuous reporter protein or another enzymatic reaction. Similar to before, screening outputs can be fluorescent, colorimetric or growth-dependent. If no detectable phenotype can be achieved, small molecules can be quantified using analytical methods, such as LC-MS or GC-MS[86].

Table 7: Screening approaches and their requirements and capacity

Method	Requirements	Capacity
Cell survival	Growth-coupled selection system	10^8 – 10^{10} per day
FACS	Fluorescent phenotype	< 10^9 per day
Colony picking	Fluorescent/ colourimetric phenotype	< 10^5 per day
Enzyme assay	Target protein performs an enzymatic reaction, which results in an easily detectable phenotype	< 10^5 per day
Analytical methods	Chromatography standards must be available	< 10^3 per day

Expression of a reporter with a conspicuous phenotype is mediated by transcription factors (Figure 21A) or riboswitches (Figure 21B), which are activated upon binding of the small molecule product. Transcription factors act on transcriptional level by binding to a specific DNA sequence close to the promoter[171]. They are described in more detail in section 2.1.1. In comparison, riboswitches act mostly on translational level. The core of the riboswitch contains a so-called aptamer, which binds the small molecule. Upon binding the ligand, riboswitches induce changes in the mRNA folding, which controls translation, yet some control on transcriptional level is also possible[583–585].

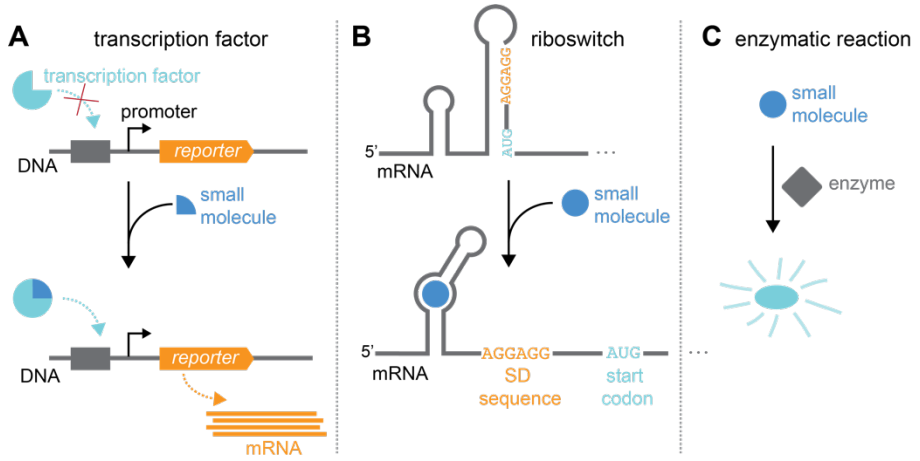


Figure 21: Inconspicuous small molecules can be converted into a detectable output. (A) Small molecules can bind to a transcription factor, which thereupon activates expression of a conspicuous reporter gene. (B) A small molecule can bind to the aptamer of a riboswitch, which thereupon induces conformational changes in the mRNA structure, resulting in the translation of a conspicuous reporter gene. (C) Inconspicuous small molecules can be converted into conspicuous small molecules in an enzymatic reaction.

However, the development of a specific sensor is time-consuming, as no fully generalizable methods to develop such sensors are available yet[582,586]. Enzymatic reactions that convert the inconspicuous small molecule into a conspicuous phenotype are rare but have been applied, e.g. in the detection of L-tyrosine which can be converted to the pigment melanin[587] (Figure 21C). If a

product is secreted into the medium, a chemical reaction with a supplemented compound can result in a detectable phenotype.

Screening capacity is generally the major bottleneck in many experiments. The development of new screening approaches is not only worthwhile for protein production optimization but is also beneficial for general metabolic engineering. The test phase of the Design-Build-Test cycle can be accelerated through the use of advanced screening methods.

Concluding remarks and future perspectives

The development of economically feasible microbial cell factories for cost efficient production of chemicals, pharmaceuticals and proteins is still a challenging task. Among others, production and screening of pathway enzymes present major bottlenecks. While the final goal might be different, here, cell factory development and recombinant protein production face a common problem. Both disciplines are confronted with low protein yields caused by insufficient expression, protein toxicity and insolubility. To overcome these challenges, different strategies can be pursued.

In this thesis we demonstrate the use of novel and versatile synthetic biology tools that improve protein production in Gram-negative and Gram-positive bacterial hosts.

In order to facilitate screening of expression libraries, we developed a set of tuneable antibiotic resistance devices that provide a simple and versatile growth-based assay for the selection of high-expressing variants in libraries with an unprecedented level of diversity. In Paper 1, we used this powerful approach to double the yield of an antibody-like protein, which was already produced at industrial titres and obtained a 65-fold increase in the yield of another antibody-like protein in *E. coli*.

To provide new tools that facilitate the use of the two Gram-positive hosts *B. subtilis* and *L. lactis* in protein production projects, we established an expression library-based approach for protein production optimization. Hereafter, we implemented a tuneable antibiotic resistance device in *B. subtilis* and *L. lactis* to select for high expressing variants. In Paper 3, the combined effort of both tools resulted in industrial protein titres with 2 g//L culture in *B. subtilis*.

Because the simple exchange of an antibiotic resistance gene or a reduction in copy number can be sufficient to increase protein yields, we developed the SEVA-Linker series, a system that allows the fast and easy exchange of plasmid backbones. With this it is possible to combine six antibiotic resistance genes and five origins of replication, resulting in 30 different plasmid backbones. The SEVA Linker series manifested a 430-fold difference in protein production of a difficult-to-express membrane protein (Paper 4).

Overall, the tools developed in this thesis facilitate the design, construction, screening and selection of gene expression constructs in many different protein production optimization approaches, but further development is still required. As an example, the tuneable antibiotic resistance devices are indicative of target gene expression levels but fail to give information about protein folding. We are considering an extension of the system for such purposes.

With advances in systems biology, researchers have uncovered a long list of unknown enzymes, which require suitable expression tools to explore their structure and function. The tools for protein production optimization developed herein may support this process. The new information that those enzymes will provide might help to understand metabolic regulation in more detail and to uncover new targets for the treatment of diseases. As a result of the former, the construction of cell factories will be accelerated and, as a result of the latter, new medication can be developed.

In general, there is a lack of adequate screening methods that keep up with the capacity and speed of parts assembly. This means that the unprecedented level of diversity that many synthetic biology approaches implement, can often not be covered in screenings. We hope that the tuneable antibiotic resistance devices, with the help of an available protocol for their use (Paper 2), will become a fundamental part of many protein production optimization processes due to being universally applicable, simple and inexpensive to use.

References Chapter 1-3

1. Irion, W. W. & Neuwirth, O. S. Oil Refining. *Ullmann's Encycl. Ind. Chem.* (2000). doi:10.1002/14356007.a18_051
2. IEA. Key World Energy Statistics 2017. (2017). doi:http://dx.doi.org/10.1787/key_energ_stat-2017-en
3. United Nations. Transforming our world: the 2030 Agenda for Sustainable Development. *Gen. Assem. 70 Sess.* **16301,** 1–35 (2015).
4. World Commission on Environment and Development. *Our Common Future*. (Oxford University Press, 1987). doi:10.1080/07488008808408783
5. Swisher, S. Sustainable Production: Definiton, Comparision, and Application. *Park Place Econ.* **14,** (2006).
6. Gavrilescu, M. & Chisti, Y. Biotechnology - A sustainable alternative for chemical industry. *Biotechnol. Adv.* **23,** 471–499 (2005).
7. Ferrer-Miralles, N., Domingo-Espín, J., Corchero, J. L., Vázquez, E. & Villaverde, A. Microbial factories for recombinant pharmaceuticals. *Microb. Cell Fact.* **8,** 17 (2009).
8. Paddon, C. J. & Keasling, J. D. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.* **12,** 355–367 (2014).
9. Galanie, S., Thodey, K., Trenchard, I. J., Interrante, M. F. & Smolke, C. D. Complete biosynthesis of opioids in yeast. *Science* **349,** 1095–1100 (2015).
10. Overton, T. W. Recombinant protein production in bacterial hosts. *Drug Discov. Today* **19,** 590–601 (2014).
11. Wang, C., Pfleger, B. F. & Kim, S. W. Reassessing Escherichia coli as a cell factory for biofuel production. *Curr. Opin. Biotechnol.* **45,** 92–103 (2017).
12. van Dijl, J. M. & Hecker, M. Bacillus subtilis: from soil bacterium to super-secreting cell factory. *Microb. Cell Fact.* **12,** 3 (2013).
13. Song, A. A.-L., In, L. L. A., Lim, S. H. E. & Rahim, R. A. A review on Lactococcus lactis: from food to factory. *Microb. Cell Fact.* **16,** 55 (2017).
14. Sauer, M. & Mattanovich, D. Construction of microbial cell factories for industrial bioprocesses. *J. Chem. Technol. Biotechnol.* **87,** 445–450 (2012).
15. Bailey, J. Toward a science of metabolic engineering. *Science* **252,** 1668–1675 (1991).
16. Weizmann, C. Production of acetone and alcohol by bacterialogical processes. **US1315585,** (1919).
17. Goeddel, D. V *et al.* Expression in Escherichia coli of chemically synthesized genes for human insulin. *Proc Natl Acad Sci U S A* **76,** 106–110 (1979).
18. Johnson, I. S. Human Insulin from Recombinant DNA Technology. *Science* **219,** 632–637 (1983).
19. Sanchez-Garcia, L. *et al.* Recombinant pharmaceuticals from microbial cells: a 2015 update. *Microb. Cell Fact.* **15,** 33 (2016).
20. Berg, C., Nassr, R. & Pang, K. The evolution of biotech. *Nat. Rev. Drug Discov.* **1,** 845–846 (2002).
21. Nagle, T., Berg, C., Nassr, R. & Pang, K. Outlook: The further evolution of biotech. *Nat. Rev. Drug Discov.* **2,** 75–79 (2003).
22. Nevin, K. P. *et al.* Electrosynthesis of organic compounds from carbon dioxide is catalyzed by a diversity of acetogenic microorganisms. *Appl. Environ. Microbiol.* **77,** 2882–2886 (2011).
23. Peralta-Yahya, P. P., Zhang, F., del Cardayre, S. B. & Keasling, J. D. Microbial engineering for the production of advanced biofuels. *Nature* **488,** 320–328 (2012).
24. Young, E. & Alper, H. Synthetic biology: Tools to design, build, and optimize cellular processes. *J. Biomed. Biotechnol.* **2010,** (2010).
25. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3,** 318–356 (1961).
26. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. Construction of Biologically Functional Bacterial Plasmids In Vitro. *Proc. Natl. Acad. Sci.* **70,** 3240–3244 (1973).
27. Jackson, D. A., Symons, R. H. & Berg, P. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **69,** 2904–2909 (1972).
28. Nathans, D. & Smith, H. O. Restriction endonucleases in the analysis and restructuring of dna molecules. *Annu. Rev. Biochem.* **44,** 273–293 (1975).

29. Saiki, R. *et al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239,** 487–491 (1988).
30. Itakura, K. *et al.* Expression in Escherichia coli of a Chemically Synthesized Gene for the Hormone Somatostatin. *Science* **198,** 1056–1063 (1977).
31. Heather, J. M. & Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **107,** 1–8 (2016).
32. Goffeau, A. A. *et al.* Life with 6000 Genes. *Science* **274,** 546+563-567 (1996).
33. Blattner, F. R. *et al.* The Complete Genome Sequence of Escherichia coli K-12 . *Science* **277,** 1453–1462 (1997).
34. Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nat. Rev. Microbiol.* **12,** 381–390 (2014).
35. Westerhoff, H. V & Palsson, B. O. The evolution of molecular biology into systems biology. *Nat. Biotechnol.* **22,** 1249–1252 (2004).
36. Jeong, H. & Albert, R. The large-scale organization of metabolic networks. **760,** 651–654 (2000).
37. Ideker, T. *et al.* Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network. *Science* **292,** 929–934 (2001).
38. Chen, R. Bacterial expression systems for recombinant protein production: E. coli and beyond. *Biotechnol. Adv.* **30,** 1102–1107 (2012).
39. Collins, J. J., Gardner, T. S. & Cantor, C. R. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403,** 339–342 (2000).
40. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403,** 335–338 (2000).
41. Shetty, R. P., Endy, D. & Knight, T. F. Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* **2,** 5 (2008).
42. Pasotti, L., Politi, N., Zucca, S., Cusella de Angelis, M. G. & Magni, P. Bottom-up engineering of biological systems through standard bricks: A modularity study on basic parts and devices. *PLoS One* **7,** 1–10 (2012).
43. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3,** (2008).
44. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).
45. Nour-Eldin, H. H., Hansen, B. G., Nørholm, M. H. H., Jensen, J. K. & Halkier, B. A. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res.* **34,** (2006).
46. Geu-Flores, F., Nour-Eldin, H. H., Nielsen, M. T. & Halkier, B. A. USER fusion: A rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res.* **35,** (2007).
47. Ferber, D. Microbes Made to Order. *Science* **303,** 158–162 (2004).
48. Liu, R., Bassalo, M. C., Zeitoun, R. I. & Gill, R. T. Genome scale engineering techniques for metabolic engineering. *Metab. Eng.* **32,** 143–154 (2015).
49. Carr, P. a & Church, G. M. Genome engineering. *Nat. Biotechnol.* **27,** 1151–1162 (2009).
50. Gibson, D. G. *et al.* Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma genitalium Genome. *Science* **319,** 1215–1220 (2008).
51. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329,** 52–56 (2010).
52. Pósfai, G., Umenhoffer, K., Kolisnychenko, V., Stahl, B. & Sharma, S. S. Emergent Properties of. *Science* **312,** 1044–1047 (2006).
53. Zhu, D. *et al.* Enhanced heterologous protein productivity by genome reduction in Lactococcus lactis NZ9000. *Microb. Cell Fact.* **16,** 1 (2017).
54. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–898 (2009).
55. Murphy, K. C. Use of bacteriophage λ recombination functions to promote gene replacement in Escherichia coli. *J. Bacteriol.* **180,** 2063–2071 (1998).
56. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 6640–6645 (2000).

57. Santos, C. N. S., Regitsky, D. D. & Yoshikuni, Y. Implementation of stable and complex biological systems through recombinase-assisted genome engineering. *Nat. Commun.* **4,** 1–10 (2013).
58. Kuhlman, T. E. & Cox, E. C. Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res.* **38,** (2010).
59. Carroll, D. Genome engineering with zinc-finger nucleases. *Genetics* **188,** 773–782 (2011).
60. Joung, J. K. & Sander, J. D. TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* **14,** 49–55 (2012).
61. Moehle, E. A. *et al.* Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc. Natl. Acad. Sci.* **104,** 3055–3060 (2007).
62. Urnov, F. D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435,** 646–651 (2005).
63. Liu, Z., Liang, Y., Ang, E. L. & Zhao, H. A New Era of Genome Integration - Simply Cut and Paste! *ACS Synth. Biol.* **6,** 601–609 (2017).
64. Charpentier, E. & Marraffini, L. A. Harnessing CRISPR-Cas9 immunity for genetic engineering. *Curr. Opin. Microbiol.* **19,** 114–119 (2014).
65. Marraffini, L. A. CRISPR-Cas immunity in prokaryotes. *Nature* **526,** 55–61 (2015).
66. Ronda, C., Pedersen, L. E., Sommer, M. O. A. & Nielsen, A. T. CRMAGE: CRISPR Optimized MAGE Recombineering. *Sci. Rep.* **6,** 19452 (2016).
67. Li, Y. *et al.* Metabolic engineering of Escherichia coli using CRISPR–Cas9 meditated genome editing. *Metab. Eng.* **31,** 13–21 (2015).
68. Pyne, M. E., Moo-Young, M., Chung, D. A. & Chou, C. P. Coupling the CRISPR/Cas9 system with lambda red recombineering enables simplified chromosomal gene replacement in Escherichia coli. *Appl. Environ. Microbiol.* **81,** 5103–5114 (2015).
69. Umenhoffer, K. *et al.* Genome-Wide Abolishment of Mobile Genetic Elements Using Genome Shuffling and CRISPR/Cas-Assisted MAGE Allows the Efficient Stabilization of a Bacterial Chassis. *ACS Synth. Biol.* **6,** 1471–1483 (2017).
70. Pines, G., Freed, E. F., Winkler, J. D. & Gill, R. T. Bacterial Recombineering: Genome Engineering via Phage-Based Homologous Recombination. *ACS Synth. Biol.* **4,** 1176–1185 (2015).
71. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25,** 1203–1210 (1997).
72. Folliard, T. *et al.* A Synthetic Recombinase-Based Feedback Loop Results in Robust Expression. *ACS Synth. Biol.* **6,** 1663–1671 (2017).
73. Klein-Marcuschamer, D., Santos, C. N. S., Yu, H. & Stephanopoulos, G. Mutagenesis of the bacterial RNA polymerase alpha subunit for improvement of complex phenotypes. *Appl. Environ. Microbiol.* **75,** 2705–2711 (2009).
74. Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F. & Voigt, C. A. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res.* **40,** 8773–8781 (2012).
75. Zhao, Y. *et al.* Visualized and precise design of artificial small RNAs for regulating T7 RNA polymerase and enhancing recombinant protein folding in Escherichia coli. *Synth. Syst. Biotechnol.* **1,** 265–270 (2016).
76. Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. A. & Gill, R. T. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat. Biotechnol.* **28,** 856–62 (2010).
77. Noren, C., Anthony-Cahill, S., Griffith, M. & Schultz, P. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244,** 182–188 (1989).
78. Völler, J. S. & Budisa, N. Coupling genetic code expansion and metabolic engineering for synthetic cells. *Curr. Opin. Biotechnol.* **48,** 1–7 (2017).
79. Li, X. & Liu, C. C. Biological applications of expanded genetic codes. *ChemBioChem* **15,** 2335–2341 (2014).
80. Liu, C. C. & Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **79,** 413–444 (2010).
81. Liu, C. C., Qi, L., Yanofsky, C. & Arkin, A. P. Regulation of transcription by unnatural amino acids. *Nat. Biotechnol.* **29,** 164–168 (2011).
82. Rackham, O. & Chin, J. W. A network of orthogonal ribosome·mRNA pairs. *Nat. Chem. Biol.* **1,** 159–166 (2005).
83. Liu, Y., Kim, D. S. & Jewett, M. C. Repurposing ribosomes for synthetic biology. *Curr. Opin. Chem. Biol.* **40,**

87–94 (2017).

84. Nielsen, J. & Keasling, J. D. Engineering Cellular Metabolism. *Cell* **164,** 1185–1197 (2016).
85. Liu, M. *et al.* Metabolic engineering of Escherichia coli to improve recombinant protein production. *Appl. Microbiol. Biotechnol.* **99,** 10367–10377 (2015).
86. Petzold, C. J., Chan, L. J. G., Nhan, M. & Adams, P. D. Analytics for Metabolic Engineering. *Front. Bioeng. Biotechnol.* **3,** 1–11 (2015).
87. Alper, H. & Stephanopoulos, G. Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nat. Rev. Microbiol.* **7,** 715–723 (2009).
88. Pfleger, B. F. & Prather, K. L. J. Biological synthesis unbounded? *Nat. Biotechnol.* **33,** 1148–1149 (2015).
89. Karp, P. D. The EcoCyc and MetaCyc databases. *Nucleic Acids Res.* **28,** 56–59 (2000).
90. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28,** 27–30 (2000).
91. Westfall, P. J. *et al.* Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc. Natl. Acad. Sci.* **109,** E111–E118 (2012).
92. Ro, D.-K. *et al.* Induction of multiple pleiotropic drug resistance genes in yeast engineered to produce an increased level of anti-malarial drug precursor, artemisinic acid. *BMC Biotechnol.* **8,** 83 (2008).
93. Paddon, C. J. *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496,** 528–532 (2013).
94. Campodonico, M. A., Andrews, B. A., Asenjo, J. A., Palsson, B. O. & Feist, A. M. Generation of an atlas for commodity chemical production in Escherichia coli and a novel pathway prediction algorithm, GEM-Path. *Metab. Eng.* **25,** 140–158 (2014).
95. Carbonell, P., Parutto, P., Herisson, J., Pandit, S. B. & Faulon, J. L. XTMS: Pathway design in an eXTended metabolic space. *Nucleic Acids Res.* **42,** 389–394 (2014).
96. Hadadi, N. & Hatzimanikatis, V. Design of computational retrobiosynthesis tools for the design of de novo synthetic pathways. *Curr. Opin. Chem. Biol.* **28,** 99–104 (2015).
97. Gustafsson, C. *et al.* Engineering genes for predictable protein expression. *Protein Expr. Purif.* **83,** 37–46 (2012).
98. McCloskey, D., Palsson, B. O. & Feist, A. M. Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. *Mol. Syst. Biol.* **9,** 661–661 (2014).
99. Burgard, A. P., Pharkya, P. & Maranas, C. D. OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. *Biotechnol. Bioeng.* **84,** 647–657 (2003).
100. O'Brien, E. J. & Palsson, B. O. Computing the functional proteome: Recent progress and future prospects for genome-scale models. *Curr. Opin. Biotechnol.* **34,** 125–134 (2015).
101. King, Z. A., O'Brien, E. J., Feist, A. M. & Palsson, B. O. Literature mining supports a next-generation modeling approach to predict cellular byproduct secretion. *Metab. Eng.* **39,** 220–227 (2017).
102. Rogers, J. K. & Church, G. M. Multiplexed Engineering in Biology. *Trends Biotechnol.* **34,** 198–206 (2016).
103. Hsu, P. D., Lander, E. S. & Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157,** 1262–1278 (2014).
104. Redden, H. & Alper, H. S. The development and characterization of synthetic minimal yeast promoters. *Nat. Commun.* **6,** 7810 (2015).
105. Peter Ruhdal Jensen, and K. H. The Sequence of Spacers between the Consensus Sequences.pdf. *Appl. enviromental Microbiol.* **64,** 82–87 (1998).
106. Guiziou, S. *et al.* A part toolbox to tune genetic expression in *Bacillus subtilis*. *Nucleic Acids Res.* **44,** gkw624 (2016).
107. Mirzadeh, K. *et al.* Enhanced Protein Production in *Escherichia coli* by Optimization of Cloning Scars at the Vector–Coding Sequence Junction. *ACS Synth. Biol.* **4,** 959–965 (2015).
108. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27,** 946–950 (2009).
109. Pfleger, B. F., Pitera, D. J., Smolke, C. D. & Keasling, J. D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.* **24,** 1027–32 (2006).
110. Smolke, C. D., Carrier, T. a & Keasling, J. D. Coordinated, Differential Expression of Two Genes through Directed mRNA Cleavage and Stabilization by Secondary Structures. *Appl. Environ. Microbiol.* **66,** 5399–

5405 (2000).

111. Rogers, J. K., Taylor, N. D. & Church, G. M. Biosensor-based engineering of biosynthetic pathways. *Curr. Opin. Biotechnol.* **42,** 84–91 (2016).
112. Saladi, S. M., Javed, N., Mueller, A. & Clemons, W. M. Decoding sequence-level information to predict membrane protein expression. *BioRxiv* 98673 (2017). doi:10.1101/098673
113. Dragosits, M. & Mattanovich, D. Adaptive laboratory evolution – principles and applications for biotechnology. *Microb. Cell Fact.* **12,** (2013).
114. Fong, S. S., Joyce, A. R. & Palsson, B. Ø. Parallel adaptive evolution cultures of Escherichia coli lead to convergent growth phenotypes with different gene expression states. *Genome Res.* **15,** 1365–1372 (2005).
115. Conrad, T. M., Lewis, N. E. & Palsson, B. O. Microbial laboratory evolution in the era of genome-scale science. *Mol. Syst. Biol.* **7,** 509–509 (2014).
116. Ibarra, R. U., Edwards, J. S. & Palsson, B. O. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **420,** 186–189 (2002).
117. Hong, K.-K., Vongsangnak, W., Vemuri, G. N. & Nielsen, J. Unravelling evolutionary strategies of yeast for improving galactose utilization through integrated systems level analysis. *Proc. Natl. Acad. Sci.* **108,** 12179–12184 (2011).
118. Kuyper, M. *et al.* Evolutionary engineering of mixed-sugar utilization by a xylose-fermenting Saccharomyces cerevisiae strain. *FEMS Yeast Res.* **5,** 925–934 (2005).
119. Sandberg, T. E. *et al.* Evolution of Escherichia coli to 42 °C and subsequent genetic engineering reveals adaptive mechanisms and novel mutations. *Mol. Biol. Evol.* **31,** 2647–2662 (2014).
120. Riehle, M. M., Bennett, A. F., Lenski, R. E. & Long, A. D. Evolutionary changes in heat-inducible gene expression in lines of *Escherichia coli* adapted to high temperature. *Physiol. Genomics* **14,** 47–58 (2003).
121. Stoebel, D. M., Hokamp, K., Last, M. S. & Dorman, C. J. Compensatory evolution of gene regulation in response to stress by Escherichia coli lacking RpoS. *PLoS Genet.* **5,** 1–9 (2009).
122. Horinouchi, T. *et al.* Transcriptome analysis of parallel-evolved Escherichia coli strains under ethanol stress. *BMC Genomics* **11,** 579 (2010).
123. Goodarzi, H. *et al.* Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in E. coli. *Mol. Syst. Biol.* **6,** 1–12 (2010).
124. Minty, J. J. *et al.* Evolution combined with genomic study elucidates genetic bases of isobutanol tolerance in Escherichia coli. *Microb. Cell Fact.* **10,** 18 (2011).
125. Atsumi, S. *et al.* Evolution, genomic analysis, and reconstruction of isobutanol tolerance in Escherichia coli. *Mol. Syst. Biol.* **6,** 1–11 (2010).
126. Mundhada, H., Schneider, K., Christensen, H. B. & Nielsen, A. T. Engineering of high yield production of L-serine in Escherichia coli. *Biotechnol. Bioeng.* **113,** 807–816 (2016).
127. Mundhada, H. *et al.* Increased production of L-serine in Escherichia coli through Adaptive Laboratory Evolution. *Metab. Eng.* **39,** 141–150 (2017).
128. Baeshen, M. N. *et al.* Production of biopharmaceuticals in E. coli: current scenario and future perspectives. *J. microbiolgy Biotechnol.* **25,** 953–962 (2015).
129. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli : advances and challenges. *Front. Microbiol.* **5,** 1–17 (2014).
130. Choi, J. H. & Lee, S. Y. Secretory and extracellular production of recombinant proteins using Escherichia coli. *Appl. Microbiol. Biotechnol.* **64,** 625–635 (2004).
131. Higgins, D. & Dworkin, J. Recent progress in Bacillus subtilis sporulation. *FEMS Microbiol. Rev.* **36,** 131–148 (2012).
132. Harwood, C. R. & Cranenburgh, R. Bacillus protein secretion: an unfolding story. *Trends Microbiol.* **16,** 73–79 (2008).
133. Jakobs, M. & Meinhardt, F. What renders Bacilli genetically competent? A gaze beyond the model organism. *Appl. Microbiol. Biotechnol.* **99,** 1557–1570 (2015).
134. Kunst, F. & al., et. The complete genome sequence of the gram-positive bacterium Bacillus subtilis. *Nature* **390,** 249–256 (1997).
135. Westbrook, A. W., Moo-Young, M. & Chou, C. P. Development of a CRISPR-Cas9 toolkit for comprehensive engineering of Bacillus subtilis. *Appl. Environ. Microbiol.* **82,** 4876–4895 (2016).
136. Song, Y., Nikoloff, J. M. & Zhang, D. Improving protein production on the level of regulation of both

expression and secretion pathways in *Bacillus subtilis*. *J. Microbiol. Biotechnol.* **25,** 963–977 (2015).

137. D'Souza, R., Raj Pandeya, D. & Hong, S.-T. Review: Lactococcus Lactis: An efficient Gram positive cell factory for the production and secretion of recombinant protein. *Biomed. Res.* **23,** 1–7 (2012).
138. Kunji, E. R. S. *et al.* Eukaryotic membrane protein overproduction in Lactococcus lactis. *Curr. Opin. Biotechnol.* **16,** 546–551 (2005).
139. Le Loir, Y. *et al.* Protein secretion in Lactococcus lactis : an efficient way to increase the overall heterologous protein production. *Microb. Cell Fact.* **4,** 2 (2005).
140. Morello, E. *et al.* Lactococcus lactis, an efficient cell factory for recombinant protein production and secretion. *J. Mol. Microbiol. Biotechnol.* **14,** 48–58 (2008).
141. Dudnik, A. *et al.* BacHBerry: BACterial Hosts for production of Bioactive phenolics from bERRY fruits. *Phytochem. Rev.* 1–36 (2017). doi:10.1007/s11101-017-9532-2
142. Song, A. A. L., Abdullah, J. O., Abdullah, M. P., Shafee, N. & Rahim, R. A. Functional expression of an orchid fragrance gene in Lactococcus lactis. *Int. J. Mol. Sci.* **13,** 1582–1597 (2012).
143. Martinez-Cuesta, M. C., Gasson, M. J. & Narbad, A. Heterologous expression of the plant coumarate: CoA ligase in Lactococcus lactis. *Lett. Appl. Microbiol.* **40,** 44–49 (2005).
144. Kavšček, M., Stražar, M., Curk, T., Natter, K. & Petrovič, U. Yeast as a cell factory: current state and perspectives. *Microb. Cell Fact.* **14,** 94 (2015).
145. Li, M. & Borodina, I. Application of synthetic biology for production of chemicals in yeast Saccharomyces cerevisiae. *FEMS Yeast Res.* **15,** 1–12 (2015).
146. Nevoigt, E. Progress in Metabolic Engineering of Saccharomyces cerevisiae. *Microbiol. Mol. Biol. Rev.* **72,** 379–412 (2008).
147. Wang, G., Huang, M. & Nielsen, J. Exploring the potential of Saccharomyces cerevisiae for biopharmaceutical protein production. *Curr. Opin. Biotechnol.* **48,** 77–84 (2017).
148. Fernández, F. J. & Vega, M. C. Technologies to keep an eye on: Alternative hosts for protein production in structural biology. *Curr. Opin. Struct. Biol.* **23,** 365–373 (2013).
149. Lubertozzi, D. & Keasling, J. D. Developing Aspergillus as a host for heterologous expression. *Biotechnol. Adv.* **27,** 53–75 (2009).
150. Anné, J., Maldonado, B., Van Impe, J., Van Mellaert, L. & Bernaerts, K. Recombinant protein production and streptomycetes. *J. Biotechnol.* **158,** 159–167 (2012).
151. Nelson, K. E. *et al.* Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440. *Environ. Microbiol.* **4,** 799–808 (2002).
152. Silva-Rocha, R. *et al.* The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res.* **41,** D666–D675 (2013).
153. Martínez-Garćía, E., Aparicio, T., Goñi-Moreno, A., Fraile, S. & De Lorenzo, V. SEVA 2.0: An update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. *Nucleic Acids Res.* **43,** D1183–D1189 (2015).
154. Aparicio, T., Jensen, S. I., Nielsen, A. T., de Lorenzo, V. & Martínez-García, E. The Ssr protein (T1E_1405) from Pseudomonas putida DOT-T1E enables oligonucleotide-based recombineering in platform strain P. putida EM42. *Biotechnol. J.* **11,** 1309–1319 (2016).
155. Martínez-García, E., Calles, B., Arévalo-Rodríguez, M. & de Lorenzo, V. pBAM1: an all-synthetic genetic tool for analysis and construction of complex bacterial phenotypes. *BMC Microbiol.* **11,** 38 (2011).
156. Gomes, N. C. M., Kosheleva, I. A., Abraham, W. R. & Smalla, K. Effects of the inoculant strain Pseudomonas putida KT2442 (pNF142) and of naphthalene contamination on the soil bacterial community. *FEMS Microbiol. Ecol.* **54,** 21–33 (2005).
157. Ward, P. G., Goff, M., Donner, M., Kaminsky, W. & O'Connor, K. E. A two step chemo-biotechnological conversion of polystyrene to a biodegradable thermoplastic. *Environ. Sci. Technol.* **40,** 2433–2437 (2006).
158. Jeon, Y. H. *et al.* Solution Structure of the Activator Contact Domain of the RNA Polymerase a Subunit. *Science* **270,** 1495–1497 (1995).
159. Zhang, G. & Darst, S. A. Structure of the Escherichia coli RNA Polymerase alpha Subunit Amino-Terminal Domain. *Science* **281,** 262–266 (1998).
160. Murakami, K. S. & Darst, S. A. Bacterial RNA polymerases: The wholo story. *Curr. Opin. Struct. Biol.* **13,** 31–39 (2003).
161. Gruber, T. M. & Gross, C. A. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription

Space. *Annu. Rev. Microbiol.* **57,** 441–466 (2003).

162. Hawley, D. K. & McClure, W. R. Compilation and analysis of Escherichia coli promoter DNA sequences. *Nucleic Acids Res.* **11,** 2237–55 (1983).
163. Gilman, J. & Love, J. Synthetic promoter design for new microbial chassis. *Biochem. Soc. Trans.* **44,** 731–737 (2016).
164. Hengge-Aronis, R. Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase. *Microbiol. Mol. Biol. Rev.* **66,** 373–95, table of contents (2002).
165. Notley-McRobb, L., King, T. & Ferenci, T. rpoS mutations and loss of general stress resistance in Escherichia coli populations as a consequence of conflict between competing stress responses. *J. Bacteriol.* **184,** 806–811 (2002).
166. Conrad, T. M. *et al.* RNA polymerase mutants found through adaptive evolution reprogram Escherichia coli for optimal growth in minimal media. *Proc. Natl. Acad. Sci.* **107,** 20500–20505 (2010).
167. Arsène, F. & Tomoyasu, T. The heat shock response of Escherichia coli. *Int. J. Food Microbiol.* **55,** 3–9 (2000).
168. Alba, B. M. & Gross, C. A. Regulation of the Escherichia coli sigma(E)-dependent envelope stress response. *Mol. Microbiol.* **52,** 613–619 (2004).
169. Bianchi, a a & Baneyx, F. Hyperosmotic shock induces the sigma32 and sigmaE stress regulons of Escherichia coli. *Mol. Microbiol.* **34,** 1029–38 (1999).
170. Pérez-Rueda, E. & Collado-Vides, J. The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. *Nucleic Acids Res.* **28,** 1838–47 (2000).
171. Babu, M. M. & Teichmann, S. A. Functional determinants of transcription factors in *Escherichia coli*: Protein families and binding sites. *Trends Genet.* **19,** 75–79 (2003).
172. Babu, M. M. & Teichmann, S. A. Evolution of transcription factors and the gene regulatory network in Escherichia coli. *Nucleic Acids Res.* **31,** 1234–1244 (2003).
173. Mascher, T., Helmann, J. D. & Unden, G. Stimulus Perception in Bacterial Signal-Transducing Histidine Kinases. *Microbiol. Mol. Biol. Rev.* **70,** 910–938 (2006).
174. Proshkin, S., Rahmouni, A. R., Mironov, A. & Nudler, E. Cooperation Between Translating Ribosomes and RNA Polymerase in Transcription Elongation. *Science* **328,** 504–508 (2010).
175. Hippel, P. H. Von. An Integrated Model of the Transcription and Editing. *Science* **281,** 660–666 (1998).
176. Gusarov, I. & Nudler, E. The mechanism of intrinsic transcription termination. *Mol. Cell* **3,** 495–504 (1999).
177. Chen, Y.-J. *et al.* Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods* **10,** 659–664 (2013).
178. Ciampi, M. S. Rho-dependent terminators and transcription termination. *Microbiology* **152,** 2515–2528 (2006).
179. Epshtein, V., Dutta, D., Wade, J. & Nudler, E. An allosteric mechanism of Rho-dependent transcription termination. *Nature* **463,** 245–249 (2010).
180. Ridgway, D. *et al.* Coarse-Grained Molecular Simulation of Diffusion and Reaction Kinetics in a Crowded Virtual Cytoplasm. *Biophys. J.* **94,** 3748–3759 (2008).
181. Brandt, F. *et al.* The Native 3D Organization of Bacterial Polysomes. *Cell* **136,** 261–271 (2009).
182. Ciandrini, L., Stansfield, I. & Romano, M. C. Ribosome Traffic on mRNAs Maps to Gene Ontology: Genome-wide Quantification of Translation Initiation Rates and Polysome Size Regulation. *PLoS Comput. Biol.* **9,** (2013).
183. Petrelli, D. *et al.* Translation initiation factor IF3: two domains, five functions, one mechanism? *EMBO J.* **20,** 4560–9 (2001).
184. Gualerzi, C., Pon, C. L. & Kaji, A. Initiation factor dependent release of aminoacyl-tRNAS from complexes of 30S ribosomal subunits, synthetic polynucleotide and aminoacyl tRNA. *Biochem. Biophys. Res. Commun.* **45,** 1312–1319 (1971).
185. Shine, J. & Dalgarno, L. The 3′-Terminal Sequence of Escherichia coli 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proc. Natl. Acad. Sci.* **71,** 1342–1346 (1974).
186. Allen, G. S., Zavialov, A., Gursky, R., Ehrenberg, M. & Frank, J. The cryo-EM structure of a translation initiation complex from Escherichia coli. *Cell* **121,** 703–712 (2005).
187. Farabaugh, P. J. Sequence of the *lacI* gene. *Nature* **274,** 765–769 (1978).

188. Hecht, A. *et al.* Measurements of translation initiation from all 64 codons in E. coli. *Nucleic Acids Res.* **45,** 3615–3626 (2017).
189. Vimberg, V., Tats, A., Remm, M. & Tenson, T. Translation initiation region sequence preferences in Escherichia coli. *BMC Mol. Biol.* **8,** 100 (2007).
190. Lockwood, a H., Chakraborty, P. R. & Maitra, U. A complex between initiation factor IF2, guanosine triphosphate, and fMet-tRNA: an intermediate in initiation complex formation. *Proc. Natl. Acad. Sci. U. S. A.* **68,** 3122–6 (1971).
191. Ibba, M. & Söll, D. Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* **69,** 617–650 (2000).
192. Berg, J. M., Tymoczko, J. L. & Stryer, L. *Biochemie*. (Elsevier GmbH., 2007).
193. Ramakrishnan, V. Ribosome structure and the mechanism of translation. *Cell* **108,** 557–572 (2002).
194. Savir, Y. & Tlusty, T. The ribosome as an optimal decoder: A lesson in molecular recognition. *Cell* **153,** 471–479 (2013).
195. Rodnina, M. V, Pape, T., Fricke, R. & Wintermeyer, W. Elongation factor Tu, a GTPase triggered by codon recognition on the ribosome: mechanism and GTP consumption. *Biochem. Cell Biol.* **73,** 1221–1227 (1995).
196. Noller, H. F., Hoffarth, V. & Zimniak, L. Unusual Resistance of Peptidyl Transferase to Protein Extraction Procedures. *Science* **256,** 1416–1419 (1992).
197. Yu, H., Chan, Y. L. & Wool, I. G. The Identification of the Determinants of the Cyclic, Sequential Binding of Elongation Factors Tu and G to the Ribosome. *J. Mol. Biol.* **386,** 802–813 (2009).
198. Dinos, G., Kalpaxis, D. L., Wilson, D. N. & Nierhaus, K. H. Deacylated tRNA is released from the E site upon A site occupation but before GTP is hydrolyzed by EF-Tu. *Nucleic Acids Res.* **33,** 5291–5296 (2005).
199. Rodnina, M. V. The ribosome in action: Tuning of translational efficiency and protein folding. *Protein Sci.* **25,** 1390–1406 (2016).
200. Young, R. & Bremer, H. Polypeptide-chain-elongation rate in Escherichia coli B/r as a function of growth rate. *Biochem. J.* **160,** 185–94 (1976).
201. Takyar, S., Hickerson, R. P. & Noller, H. F. mRNA helicase activity of the ribosome. *Cell* **120,** 49–58 (2005).
202. Qu, X. *et al.* The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature* **475,** 118–121 (2011).
203. Kisselev, L. L. & Buckingham, R. H. Translational termination comes of age. *Trends Biochem. Sci.* **25,** 561–566 (2000).
204. Savelsbergh, A., Rodnina, M. V & Wintermeyer, W. Distinct functions of elongation factor G in ribosome recycling and translocation. *RNA* **15,** 772–780 (2009).
205. Hirokawa, G., Demeshkina, N., Iwakura, N., Kaji, H. & Kaji, A. The ribosome-recycling step: Consensus or controversy? *Trends Biochem. Sci.* **31,** 143–149 (2006).
206. Tian, T. & Salis, H. M. A predictive biophysical model of translational coupling to coordinate and control protein expression in bacterial operons. *Nucleic Acids Res.* **43,** 7137–7151 (2015).
207. Mehta, R. & Champney, W. S. Neomycin and paromomycin inhibit 30S ribosomal subunit assembly in Staphylococcus aureus. *Curr. Microbiol.* **47,** 237–243 (2003).
208. Wilson, D. N. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat. Rev. Microbiol.* **12,** 35–48 (2014).
209. Wilson, D. N. The A–Z of bacterial translation inhibitors. *Crit. Rev. Biochem. Mol. Biol.* **44,** 393–433 (2009).
210. Misumi, M., Nishimura, T., Komai, T. & Tanaka, N. Interaction of kanamycin and related antibiotics with the large subunit of ribosomes and the inhibition of translocation. *Biochem. Biophys. Res. Commun.* **84,** 358–365 (1978).
211. Kohanski, M. A., Dwyer, D. J. & Collins, J. J. How antibiotics kill bacteria: from targets to networks. *Nat. Rev. Microbiol.* **8,** 423–435 (2010).
212. Moazed, D. & Noller, H. F. Chloramphenicol, erythromycin, carbomycin and vernamycin B protect overlapping sites in the peptidyl transferase region of 23S ribosomal RNA. *Biochimie* **69,** 879–884 (1987).
213. Campbell, E. A. *et al.* Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* **104,** 901–912 (2001).
214. Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science* **181,** 223–230 (1973).
215. Jahn, T. R. & Radford, S. E. The Yin and Yang of protein folding. *FEBS J.* **272,** 5962–5970 (2005).
216. Hartl, F. U. & Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **16,** 574–581 (2009).

217. Kim, H. K., Choi, S. Il & Seong, B. L. 5S rRNA-assisted DnaK refolding. *Biochem. Biophys. Res. Commun.* **391,** 1177–1181 (2010).

218. Dowhan, W. & Bogdanov, M. Molecular genetic and biochemical approaches for defining lipid-dependent membrane protein folding. *Biochim. Biophys. Acta - Biomembr.* **1818,** 1097–1107 (2012).

219. Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M. & Ulrich Hartl, F. *Molecular Chaperone Functions in Protein Folding and Proteostasis*. *Annual Review of Biochemistry* **82,** (2013).

220. Hjelm, A. Optimizing membrane and secretory protein production in Gram-negative bacteria. (Stockholm University, Sweden, 2015).

221. Castanié-Cornet, M. P., Bruel, N. & Genevaux, P. Chaperone networking facilitates protein targeting to the bacterial cytoplasmic membrane. *Biochim. Biophys. Acta - Mol. Cell Res.* **1843,** 1442–1456 (2014).

222. Hebert, D. N., Chandrasekhar, K. D. & Gierasch, L. M. You Got to Know When to Hold (or Unfold) 'Em... *Mol. Cell* **48,** 3–4 (2012).

223. Calloni, G. *et al.* DnaK Functions as a Central Hub in the E. coli Chaperone Network. *Cell Rep.* **1,** 251–264 (2012).

224. Kerner, M. J. *et al.* Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. *Cell* **122,** 209–220 (2005).

225. Hayer-Hartl, M., Bracher, A. & Hartl, F. U. The GroEL-GroES Chaperonin Machine: A Nano-Cage for Protein Folding. *Trends Biochem. Sci.* **41,** 62–76 (2016).

226. Braig, K. *et al.* The crystal structure of the bacterial chaperonln GroEL at 2.8 Å. *Nature* **371,** 578–586 (1994).

227. Xu, Z., Horwich, A. L. & Sigler, P. B. The crystal structure of the asymmetric GroEL–GroES–(ADP)7 chaperonin complex. *Nature* **388,** 741–750 (1997).

228. J.Silhavy, T., Kahne, D. & Walker, S. The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* **2,** 1–17 (2010).

229. Tsirigotaki, A., De Geyter, J., Šoštaric´, N., Economou, A. & Karamanou, S. Protein export through the bacterial Sec pathway. *Nat. Rev. Microbiol.* **15,** 21–36 (2016).

230. Natale, P., Brüser, T. & Driessen, A. J. M. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-Distinct translocases and mechanisms. *Biochim. Biophys. Acta - Biomembr.* **1778,** 1735–1756 (2008).

231. Osborne, A. R., Rapoport, T. A. & van den Berg, B. Protein Translocation By the Sec61/Secy Channel. *Annu. Rev. Cell Dev. Biol.* **21,** 529–550 (2005).

232. Luirink, J., Yu, Z., Wagner, S. & de Gier, J.-W. Biogenesis of inner membrane proteins in Escherichia coli. *Biochim. Biophys. Acta - Bioenerg.* **1817,** 965–976 (2012).

233. Keenan, R. J., Freymann, D. M., Stroud, R. M. & Walter, P. The Signal Recognition Particle. *Annu. Rev. Biochem.* **70,** 755–775 (2001).

234. Estrozi, L. F., Boehringer, D., Shan, S., Ban, N. & Schaffitzel, C. Cryo-EM structure of the E. coli translating ribosome in complex with SRP and its receptor. *Nat. Struct. Mol. Biol.* **18,** 88–90 (2011).

235. Draycheva, A., Bornemann, T., Ryazanov, S., Lakomek, N. A. & Wintermeyer, W. The bacterial SRP receptor, FtsY, is activated on binding to the translocon. *Mol. Microbiol.* **102,** 152–167 (2016).

236. von Loeffelholz, O. *et al.* Ribosome-SRP-FtsY cotranslational targeting complex in the closed state. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 3943–3948 (2015).

237. Samuelson, J. C. *et al.* YidC mediates membrane protein insertion in bacteria. *Nature* **406,** 637–641 (2000).

238. Facey, S. J., Neugebauer, S. A., Krauss, S. & Kuhn, A. The Mechanosensitive Channel Protein MscL Is Targeted by the SRP to The Novel YidC Membrane Insertion Pathway of Escherichia coli. *J. Mol. Biol.* **365,** 995–1004 (2007).

239. Dalbey, R. E., Kuhn, A., Zhu, L. & Kiefer, D. The membrane insertase YidC. *Biochim. Biophys. Acta - Mol. Cell Res.* **1843,** 1489–1496 (2014).

240. van der Laan, M. *et al.* A conserved function of YidC in the biogenesis of respiratory chain complexes. *Proc. Natl. Acad. Sci.* **100,** 5801–5806 (2003).

241. von Heijne, G. The Signal peptide. *J Membr Biol* **115,** 195–201 (1990).

242. Steiner, D., Forrer, P., Stumpp, M. T. & Plückthun, A. Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. *Nat. Biotechnol.* **24,** 823–831 (2006).

243. Huber, D. *et al.* Use of thioredoxin as a reporter to identify a subset of Escherichia coli signal sequences that promote signal recognition particle-dependent translocation. *J. Bacteriol.* **187,** 2983–2991 (2005).

244. Kudva, R. *et al.* Protein translocation across the inner membrane of Gram-negative bacteria: The Sec and Tat dependent protein transport pathways. *Res. Microbiol.* **164,** 505–534 (2013).
245. Bornemann, T., Holtkamp, W. & Wintermeyer, W. Interplay between trigger factor and other protein biogenesis factors on the ribosome. *Nat. Commun.* **5,** 1–8 (2014).
246. Mitra, K., Frank, J. & Driessen, A. Co- and post-translational translocation through the protein-conducting channel: analogous mechanisms at work? *Nat. Struct. Mol. Biol.* **13,** 957–964 (2006).
247. Huang, Q. & Palmer, T. Signal Peptide Hydrophobicity Modulates Interaction with the Twin-Arginine Translocase. *MBio* **8,** e00909-17 (2017).
248. Weiner, J. H. *et al.* A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins. *Cell* **93,** 93–101 (1998).
249. DeLisa, M. P., Tullman, D. & Georgiou, G. Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *Proc. Natl. Acad. Sci. U S A* **100,** 6115–6120 (2003).
250. Pérez-Rodríguez, R. *et al.* An Essential Role for the DnaK Molecular Chaperone in Stabilizing Over-expressed Substrate Proteins of the Bacterial Twin-arginine Translocation Pathway. *J. Mol. Biol.* **367,** 715–730 (2007).
251. Alami, M. *et al.* Differential interactions between a twin-arginine signal peptide and its translocase in Escherichia coli. *Mol. Cell* **12,** 937–946 (2003).
252. Berks, B. C., Sargent, F. & Palmer, T. The Tat protein export pathway. *Mol Microbiol.* **35,** 260–274 (2000).
253. Mergulhao, F. J. M., Summers, D. K. & Monteiro, G. A. Recombinant protein secretion in Escherichia coli. *Biotechnol. Adv.* **23,** 177–202 (2005).
254. Götzke, H. *et al.* Identification of Putative Substrates for the Periplasmic Chaperone YfgM in *Escherichia coli* Using Quantitative Proteomics. *Mol. Cell. Proteomics* **14,** 216–226 (2015).
255. Götzke, H. *et al.* YfgM is an ancillary subunit of the secYEG translocon in Escherichia coli. *J. Biol. Chem.* **289,** 19089–19097 (2014).
256. Antonoaea, R., Fürst, M., Nishiyama, K. I. & Müller, M. The periplasmic chaperone PpiD interacts with secretory proteins exiting from the SecYEG translocon. *Biochemistry* **47,** 5649–5656 (2008).
257. Quan, S. *et al.* Genetic selection designed to stabilize proteins uncovers a chaperone called Spy. *Nat. Struct. Mol. Biol.* **18,** 262–269 (2011).
258. Schwalm, J., Mahoney, T. F., Soltes, G. R. & Silhavy, T. J. Role for Skp in LptD assembly in Escherichia coli. *J. Bacteriol.* **195,** 3734–3742 (2013).
259. Saul, F. A. *et al.* Structural and Functional Studies of FkpA from Escherichia coli, a cis/trans Peptidyl-prolyl Isomerase with Chaperone Activity. *J. Mol. Biol.* **335,** 595–608 (2004).
260. Dutton, R. J., Boyd, D., Berkmen, M. & Beckwith, J. Bacterial species exhibit diversity in their mechanisms and capacity for protein disulfide bond formation. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 11933–11938 (2008).
261. Hatahet, F., Boyd, D. & Beckwith, J. Disulfide bond formation in prokaryotes: History, diversity and design. *Biochim. Biophys. Acta - Proteins Proteomics* **1844,** 1402–1414 (2014).
262. Shouldice, S. R. *et al.* In vivo oxidative protein folding can be facilitated by oxidation-reduction cycling. *Mol. Microbiol.* **75,** 13–28 (2010).
263. Kadokura, H., Katzen, F. & Beckwith, J. Protein Disulfide Bond Formation in Prokaryotes. *Annu. Rev. Biochem.* **72,** 111–135 (2003).
264. Knowles, T. J., Scott-Tucker, A., Overduin, M. & Henderson, I. R. Membrane protein architects: the role of the BAM complex in outer membrane protein assembly. *Nat. Rev. Microbiol.* **7,** 206–214 (2009).
265. Behrens-Kneip, S. The role of SurA factor in outer membrane protein transport and virulence. *Int. J. Med. Microbiol.* **300,** 421–428 (2010).
266. Bitto, E. & McKay, D. B. Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. *Structure* **10,** 1489–1498 (2002).
267. Sklar, J. G., Wu, T., Kahne, D. & Silhavy, T. J. Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli. *Genes Dev.* **21,** 2473–2484 (2007).
268. Rondelet, A. & Condemine, G. Sura Is involved in the targeting to the outer membrane of a tat signal sequence-anchored protein. *J. Bacteriol.* **194,** 6131–6142 (2012).
269. Walton, T. A., Sandoval, C. M., Fowler, C. A., Pardi, A. & Sousa, M. C. The cavity-chaperone Skp protects its substrate from aggregation but allows independent folding of substrate domains. *Proc. Natl. Acad. Sci. U S A* **106,** 1772–1777 (2009).

270. Schiffrin, B. *et al.* Skp is a multivalent chaperone of outer-membrane proteins. *Nat. Struct. Mol. Biol.* **23,** 786–793 (2016).
271. Ge, X. *et al.* DegP primarily functions as a protease for the biogenesis of β-barrel outer membrane proteins in the Gram-negative bacterium Escherichia coli. *FEBS J.* **281,** 1226–1240 (2014).
272. Noinaj, N., Gumbart, J. C. & Buchanan, S. K. The β-barrel assembly machinery in motion. *Nat. Rev. Microbiol.* **15,** 197–204 (2017).
273. Noinaj, N., Rollauer, S. E. & Buchanan, S. K. The β-barrel membrane protein insertase machinery from Gram-negative bacteria. *Curr. Opin. Struct. Biol.* **31,** 35–42 (2015).
274. Gu, Y. *et al.* Structural basis of outer membrane protein insertion by the BAM complex. *Nature* **531,** 64–69 (2016).
275. Han, L. *et al.* Structure of the BAM complex and its implications for biogenesis of outer-membrane proteins. *Nat. Struct. Mol. Biol.* **23,** 192–196 (2016).
276. Iadanza, M. G. *et al.* Lateral opening in the intact β-barrel assembly machinery captured by cryo-EM. *Nat. Commun.* **7,** 12865 (2016).
277. Costa, T. R. D. *et al.* Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat. Rev. Microbiol.* **13,** 343–359 (2015).
278. Sørensen, H. P. & Mortensen, K. K. Advanced genetic strategies for recombinant protein expression in Escherichia coli. *J. Biotechnol.* **115,** 113–128 (2005).
279. Gronenborn, B. Overproduction of phage lambda repressor under control of the lac promotor of Escherichia coli. *Mol. Gen. Genet.* **148,** 243–250 (1976).
280. Guzman, L.-M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177,** 4121–4130 (1995).
281. Haldimann, A., Daniels, L. L. & Wanner, B. L. Use of new methods for construction of tightly regulated arabinose and rhamnose promoter fusions in studies of the Escherichia coli phosphate regulon. *J. Bacteriol.* **180,** 1277–1286 (1998).
282. Elvin, C. M. *et al.* Modified bacteriophage lambda promoter vectors for overproduction of proteins in Escherichia coli. *Gene* **87,** 123–126 (1990).
283. Davis, J. H., Rubin, A. J. & Sauer, R. T. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res.* **39,** 1131–1141 (2011).
284. Wickstrum, J. R., Santangelo, T. J. & Egan, S. M. Cyclic AMP Receptor Protein and RhaR Synergistically Activate Transcription from the L-Rhamnose-Responsive *rhaSR* promoter in *Escherichia coli*. *J. Bacteriol.* **187,** 6708–6718 (2005).
285. Via, P., Badia, J., Baldoma, L., Obradors, N. & Aguilar, J. Transcriptional regulation of the Escherichia coli rhaT gene. *Microbiology* **142,** 1833–1840 (1996).
286. Egan, S. M. & Schleif, R. F. A Regulatory Cascade in the Induction of rhaBAD. *Journal of Molecular Biology* **234,** 87–98 (1993).
287. Giacalone, M. J. *et al.* Toxic protein expression in Escherichia coli using a rhamnose-based tightly regulated and tunable promoter system. *Biotechniques* **40,** 355–364 (2006).
288. Wickstrum, J. R., Skredenske, J. M., Balasubramaniam, V., Jones, K. & Egan, S. M. The AraC/XylS family activator RhaS negatively autoregulates rhaSR expression by preventing cyclic AMP receptor protein activation. *J. Bacteriol.* **192,** 225–232 (2010).
289. Görke, B. & Stülke, J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat. Rev. Microbiol.* **6,** 613–624 (2008).
290. Kolb, A., Busby, S., Buc, H., Garges, S. & Adhya, S. Transcriptional regulation by cAMP and its receptor protein. *Annu. Rev. Biochem.* **62,** 749–795 (1993).
291. Hjelm, A. *et al.* Tailoring Escherichia coli for the L-rhamnose PBAD promoter-based production of membrane and secretory proteins. *ACS Synth. Biol.* **6,** 985–994 (2017).
292. Kelly, C. L. *et al.* Synthetic Chemical Inducers and Genetic Decoupling Enable Orthogonal Control of the rhaBAD Promoter. *ACS Synth. Biol.* **5,** 1136–1145 (2016).
293. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189,** 113–130 (1986).
294. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185,** 60–89 (1990).

295. Wanner, B. L., Kodaira, R. & Neidhardt, F. C. Physiological regulation of a decontrolled lac operon. *J. Bacteriol.* **130,** 212–222 (1977).
296. Dubendorff, J. W. & Studier, F. W. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* **219,** 45–59 (1991).
297. Iost, I., Guillerez, J. & Dreyfus, M. Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. *J. Bacteriol.* **174,** 619–622 (1992).
298. Studier, F. W. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* **219,** 37–44 (1991).
299. Stano, N. M. & Patel, S. S. T7 Lysozyme Represses T7 RNA Polymerase Transcription by Destabilizing the Open Complex during Initiation. *J. Biol. Chem.* **279,** 16136–16143 (2004).
300. Lewis, M. The lac repressor. *C. R. Biol.* **328,** 521–548 (2005).
301. Brosius, J., Erfle, M. & Storella, J. Spacing of the -10 and -35 Regions in the tac Promoter. Effect on its in vivo activity. *J. Biol. Chem.* **260,** 3539–3541 (1985).
302. de Boer, H. A., Comstock, L. J. & Vasser, M. The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proc. Natl. Acad. Sci. U S A* **80,** 21–25 (1983).
303. Megerle, J. A., Fritz, G., Gerland, U., Jung, K. & Rädler, J. O. Timing and Dynamics of Single Cell Gene Expression in the Arabinose Utilization System. *Biophys. J.* **95,** 2103–2115 (2008).
304. Siegele, D. a & Hu, J. C. Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 8168–8172 (1997).
305. Schleif, R. AraC protein, regulation of the l-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action. *FEMS Microbiol. Rev.* **34,** 779–796 (2010).
306. Khlebnikov, A., Datsenko, K. A., Skaug, T., Wanner, B. L. & Keasling, J. D. Homogeneous expression of the PBAD promoter in Escherichia coli by constitutive expression of the low-affinity high-capacity araE transporter. *Microbiology* **147,** 3241–3247 (2001).
307. Khlebnikov, A., Skaug, T. & Keasling, J. D. Modulation of gene expression from the arabinose-inducible araBAD promoter. *J. Ind. Microbiol. Biotechnol.* **29,** 34–37 (2002).
308. Muthukrishnan, A. B. *et al.* Dynamics of transcription driven by the tetA promoter, one event at a time, in live Escherichia coli cells. *Nucleic Acids Res.* **40,** 8472–8483 (2012).
309. Beck, C. F., Mutzel, R., Barbe, J. & Muller, W. A multifunctional gene (tetR) controls Tn10-encoded tetracycline resistance. *J. Bacteriol.* **150,** 633–642 (1982).
310. Hillen, W. & Berens, C. Mechansims underlying expression of Tn10 encoded tetracycline resistance. *Annu. Rev. Microbiol.* **48,** 345–69 (1994).
311. Skerra, A. Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in Escherichia coli. *Gene* **151,** 131–135 (1994).
312. Degenkolb, J., Takahashi, M., Ellestad, G. A. & Hillen, W. Structural requirements of tetracycline-Tet repressor interaction: Determination of equilibrium binding constants for tetracycline analogs with the Tet repressor. *Antimicrob. Agents Chemother.* **35,** 1591–1595 (1991).
313. Valdez-Cruz, N. a, Caspeta, L., Pérez, N. O., Ramírez, O. T. & Trujillo-Roldán, M. a. Production of recombinant proteins in E. coli by the heat inducible expression system based on the phage lambda pL and/or pR promoters. *Microb. Cell Fact.* **9,** 18 (2010).
314. Schnappinger, D. & Hillen, W. Tetracyclines: Antibiotic action, uptake, and resistance mechanisms. *Arch. Microbiol.* **165,** 359–369 (1996).
315. Kim, S. H., Cavaleiro, A. M., Rennig, M. & Nørholm, M. H. H. SEVA linkers: a versatile and automatable DNA backbone exchange standard for synthetic biology. *ACS Synth. Biol.* **5,** 1177–1181 (2016).
316. Andersson, D. I. & Levin, B. R. The biological cost of antibiotic resistance. *Curr. Opin. Microbiol.* **2,** 489–493 (1999).
317. Martínez, J. L. & Rojo, F. Metabolic regulation of antibiotic resistance. *FEMS Microbiol. Rev.* **35,** 768–789 (2011).
318. Zur Wiesch, P. S., Engelstädter, J. & Bonhoeffer, S. Compensation of fitness costs and reversibility of antibiotic resistance mutations. *Antimicrob. Agents Chemother.* **54,** 2085–2095 (2010).
319. Hägg, P., De Pohl, J. W., Abdulkarim, F. & Isaksson, L. A. A host/plasmid system that is not dependent on antibiotics and antibiotic resistance genes for stable plasmid maintenance in Escherichia coli. *J.*

Biotechnol. **111,** 17–30 (2004).

320. Peubez, I. *et al.* Antibiotic-free selection in E. coli: new considerations for optimal design and improved production. *Microb. Cell Fact.* **9,** 65 (2010).
321. Hershfield, V., Boyer, H. W., Yanofsky, C., Lovett, M. A. & Helinski, D. R. Plasmid ColEl as a molecular vehicle for cloning and amplification of DNA. *Proc. Natl. Acad. Sci. U. S. A.* **71,** 3455–9 (1974).
322. Hershfield, V., Boyer, H. W., Chow, L. & Helinski, D. R. Characterization of a Mini-ColE1 Plasmid. **126,** 447–453 (1976).
323. Kahn, M. *et al.* Plasmid Cloning Vehicles Derived from Plasmids ColE1, F, R6K, and RK2. *Methods Enzymol.* **68,** 268–280 (1979).
324. Vieira, J. & Messing, J. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19,** 259–268 (1982).
325. Lin-Chao, S., Chen, W.-T. & Wong, T.-T. High copy number of the pUC piasmid results from a Rom / Rop-suppressible point mutation In RNA II. *Mol. Microbiol.* **6,** 3385–3393 (1992).
326. Jeffrey, V. & Joachim, M. New pUC-derived cloning vectors with different selectable markers and DNA replication origins. *Gene* **100,** 189–194 (1991).
327. Short, J. M., Fernandez, J. M., Sorge, J. A. & Huse, W. D. λ ZAP: A bacteriophage λ expression vector with in vivo excision properties. *Nucleic Acids Res.* **16,** 7583–7600 (1988).
328. Bolivar, F. *et al.* Construction and characterization of new cloning vehicles - II. A multipurpose cloning system. *Gene* **2,** 95–113 (1977).
329. Balbás, P. *et al.* Plasmid vector pBR322 and its special-purpose derivatives - a review. *Gene* **50,** 3–40 (1986).
330. Chang, A. C. Y. & Cohen, S. N. Construction and characterization of amplifiable DNA cloning vectors derived from P15A cryptic plasmid. *J. Bacteriol.* **134,** 1141–1156. (1978).
331. Stoker, N. G., Fairweathe, N. F. & Spratt, B. G. Versatile low-copy-number plasmid vectors for cloning in Escherichia coli. *Gene* **18,** 335–341 (1982).
332. Hashimoto-Gotoh, T., Franklin, F. C. H., Nordheim, A. & Timmis, K. N. Specific-purpose plasmid cloning vectors I. Low copy number, temperature-sensitive, mobilization-defective pSC101-derived containment vectors. *Gene* **16,** 227–235 (1981).
333. Schweizer, H. P. Bacterial genetics: Past achievements, present state of the field, and future challenges. *Biotechniques* **44,** 633–641 (2008).
334. Antoine, R. & Locht, C. Isolation and molecular characterization of a novel broad host range plasmid from Bordetella bronchiseptica with sequence similarities to plasmids from Gram positive organisms. *Mol. Microbiol.* **6,** 1785–1799 (1992).
335. Kovach, M. E. *et al.* Four new derivatives of the broad host range cloning vector PBBR1MCS, carrying different antibiotic resistance cassettes. *Gene* **166,** 175–176 (1995).
336. Kües, U. & Stahl, U. Replication of plasmids in gram-negative bacteria. *Microbiol Rev* **53,** 491–516 (1989).
337. Scholz, P. *et al.* Complete nucleotide sequence and gene organization of the broad-host-range plasmid RSF1010. *Gene* **75,** 271–288 (1989).
338. Olsen, R. H., DeBusscher, G. & McCombie, W. R. Development of broad-host-range vectors and gene banks: Self-cloning of the Pseudomonas aeruginosa PAO chromosome. *J. Bacteriol.* **150,** 60–69 (1982).
339. Schweizer, H. P. Escherichia-Pseudomonas shuttle vectors derived from pUC18/19. *Gene* **97,** 109–112 (1991).
340. Stalker, D. M., Kolter, R. & Helinski, D. R. Nucleotide sequence of the region of an origin of replication of the antibiotic resistance plasmid R6K. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 1150–4 (1979).
341. Linares, D. M., Kok, J. & Poolman, B. Genome sequences of Lactococcus lactis MG1363 (revised) and NZ9000 and comparative physiological studies. *J. Bacteriol.* **192,** 5806–5812 (2010).
342. Zhu, D. *et al.* Isolation of strong constitutive promoters from Lactococcus lactis subsp. lactis N8. *FEMS Microbiol. Lett.* **362,** 1–6 (2015).
343. Zhou, X. X., Li, W. F., Ma, G. X. & Pan, Y. J. The nisin-controlled gene expression system: Construction, application and improvements. *Biotechnol. Adv.* **24,** 285–295 (2006).
344. De Ruyter, P. G. G. A., Kuipers, O. P., Beerthuyzen, M. M., Van Alen-Boerrigter, I. & De Vos, W. M. Functional analysis of promoters in the nisin gene cluster of Lactococcus lactis. *J. Bacteriol.* **178,** 3434–3439 (1996).
345. de Ruyter, P. G., Kuipers, O. P. & de Vos, W. M. Controlled gene expression systems for Lactococcus

lactis with the food-grade inducer nisin. *Appl. Environ. Microbiol.* **62,** 3662–3667 (1996).

346. Morello, E. *et al.* Lactococcus lactis, an efficient cell factory for recombinant protein production and secretion. *J. Mol. Microbiol. Biotechnol.* **14,** 48–58 (2007).
347. van der Vossen, J. M., van der Lelie, D. & Venema, G. Isolation and characterization of *Streptococcus cremoris* Wg2-specific promoters. *Appl. Environ. Microbiol.* **53,** 2452–2457 (1987).
348. Koivula, T., Sibakov, M. & Palva, I. Isolation and Characterization of Lactococcus lactis subsp. lactis Promoters. *Appl. Environ. Microbiol.* **57,** 333–340 (1991).
349. Jørgensen, C. M., Vrang, A. & Madsen, S. M. Recombinant protein expression in Lactococcus lactis using the P170 expression system. *FEMS Microbiol. Lett.* **351,** 170–178 (2014).
350. Nauta, A., van den Burg, B., Karsens, H., Venema, G. & Kok, J. Design of thermolabile bacteriophage repressor mutants by comparative molecular modeling. *Nat. Biotechnol.* **15,** 980–983 (1997).
351. Walker, S. A. & Klaenhammer, T. R. Molecular characterization of a phage-inducible middle promoter and its transcriptional activator from the lactococcal bacteriophage phi31. *J. Bacteriol.* **180,** 921–931 (1998).
352. Sirén, N., Salonen, K., Leisola, M. & Nyyssölä, A. A new and efficient phosphate starvation inducible expression system for Lactococcus lactis. *Appl. Microbiol. Biotechnol.* **79,** 803–810 (2008).
353. Miyoshi, A. *et al.* A xylose-inducible expression system for Lactococcus lactis. *FEMS Microbiol. Lett.* **239,** 205–212 (2004).
354. Van Rooijen, R. J., Gasson, M. J. & De Vos, W. M. Characterization of the Lactococcus lactis lactose operon promoter: Contribution of flanking sequences and lacR repressor to promoter activity. *J. Bacteriol.* **174,** 2273–2280 (1992).
355. Llull, D. & Poquet, I. New expression system tightly controlled by zinc availability in Lactococcus lactis. *Appl. Environ. Microbiol.* **70,** 5398–5406 (2004).
356. Mu, D., Montalbán-López, M., Masuda, Y. & Kuipers, O. P. Zirex: A novel zinc-regulated expression system for Lactococcus lactis. *Appl. Environ. Microbiol.* **79,** 4503–4508 (2013).
357. Lakowitz, A., Godard, T., Biedendieck, R. & Krull, R. Mini review: Recombinant production of tailored bio-pharmaceuticals in different Bacillus strains and future perspectives. *Eur. J. Pharm. Biopharm.* (2017). doi:10.1016/j.ejpb.2017.06.008
358. Song, Y. *et al.* Promoter screening from Bacillus subtilis in various conditions hunting for synthetic biology and industrial applications. *PLoS One* **11,** 1–18 (2016).
359. Lam, K. H., Chow, K. C. & Wong, W. K. Construction of an efficient *Bacillus subtilis* system for extracellular production of heterologous proteins. *J. Biotechnol.* **63,** 167–77 (1998).
360. Qin, Y. *et al.* The Highly Conserved LepA Is??a Ribosomal Elongation Factor that Back-Translocates the Ribosome. *Cell* **127,** 721–733 (2006).
361. Radeck, J. *et al.* The Bacillus BioBrick Box: generation and evaluation of essential genetic building blocks for standardized work with Bacillus subtilis. *J. Biol. Eng.* **7,** 29 (2013).
362. Jordan, S., Junker, A., Helmann, J. D. & Mascher, T. Regulation of LiaRS-dependent gene expression in Bacillus subtilis: Identification of inhibitor proteins, regulator binding sites, and target genes of a conserved cell envelope stress-sensing two-component system. *J. Bacteriol.* **188,** 5153–5166 (2006).
363. Bongers, R. S. *et al.* Development and Characterization of a Subtilin-Regulated Expression System in Bacillus subtilis : Strict Control of Gene Expression by Addition of Subtilin Development and Characterization of a Subtilin-Regulated Expression System in Bacillus subtilis : S. *Appl. Environ. Microbiol.* **71,** 8818–24 (2005).
364. Vavrová, L., Muchová, K. & Barák, I. Comparison of different Bacillus subtilis expression systems. *Res. Microbiol.* **161,** 791–797 (2010).
365. Paccez, J. D. *et al.* Stable episomal expression system under control of a stress inducible promoter enhances the immunogenicity of Bacillus subtilis as a vector for antigen delivery. *Vaccine* **24,** 2935–2943 (2006).
366. Lee, J. K., Edwards, C. W. & Hulett, F. M. Bacillus licheniformis APase I gene promoter: a strong well-regulated promoter in B. subtilis. *J. Gen. Microbiol.* **137,** 1127–33 (1991).
367. Zhang, W. W., Gao, Q. R., Yang, M. M., Liu, H. & Wang, D. Assay and characterization of an osmolarity inducible promoter newly isolated from Bacillus subtilis. *Mol. Biol. Rep.* **39,** 7347–7353 (2012).
368. Li, W. *et al.* Characterization of two temperature-inducible promoters newly isolated from B. subtilis. *Biochem. Biophys. Res. Commun.* **358,** 1148–1153 (2007).

369. Thuy Le, A. T. & Schumann, W. A novel cold-inducible expression system for Bacillus subtilis. *Protein Expr. Purif.* **53,** 264–269 (2007).

370. Bhavsar, A. P., Zhao, X. & Brown, E. D. Development and Characterization of a Xylose-Dependent System for Expression of Cloned Genes in Bacillus subtilis : Conditional Complementation of a Teichoic Acid Mutant Development and Characterization of a Xylose-Dependent System for Expression of Clone. **67,** 1–3 (2001).

371. Kim, L., Mogk, A. & Schumann, W. A xylose-inducible Bacillus subtilis integration vector and its application. *Gene* **181,** 71–76 (1996).

372. Phan, T. T. P., Nguyen, H. D. & Schumann, W. Development of a strong intracellular expression system for Bacillus subtilis by optimizing promoter elements. *J. Biotechnol.* **157,** 167–172 (2012).

373. Chen, P. T., Shaw, J. F., Chao, Y. P., David Ho, T. H. & Yu, S. M. Construction of chromosomally located T7 expression system for production of heterologous secreted proteins in bacillus subtilis. *J. Agric. Food Chem.* **58,** 5392–5399 (2010).

374. Conrad, B., Savchenko, R. S., Breves, R. & Hofemeister, J. A T7 promoter-specific, inducible protein expression system for Bacillus subtilis. *Mol. Gen. Genet.* **250,** 230–236 (1996).

375. Toymentseva, A. A., Schrecke, K., Sharipova, M. R. & Mascher, T. The LIKE system, a novel protein expression toolbox for Bacillus subtilis based on the liaI promoter. *Microb. Cell Fact.* **11,** 143 (2012).

376. Geissendörfer, M. & Hillen, W. Regulated expression of heterologous genes in Bacillus subtilis using the Tn10 encoded tet regulatory elements. *Appl. Microbiol. Biotechnol.* **33,** 657–63 (1990).

377. Mascher, T., Zimmer, S. L., Smith, T. A. & Helmann, J. D. Antibiotic-inducible promoter regulated by the cell envelope stress-sensing two-component system LiaRS of Bacillus subtilis. *Antimicrob. Agents Chemother.* **48,** 2888–2896 (2004).

378. Liu, S. L. & Du, K. Enhanced expression of an endoglucanase in Bacillus subtilis by using the sucrose-inducible sacB promoter and improved properties of the recombinant enzyme. *Protein Expr. Purif.* **83,** 164–168 (2012).

379. Morabbi Heravi, K., Wenzel, M. & Altenbuchner, J. Regulation of mtl operon promoter of Bacillus subtilis: requirements of its use in expression vectors. *Microb. Cell Fact.* **10,** 83 (2011).

380. Ming, Y. M., Wei, Z. W., Lin, C. Y. & Sheng, G. Y. Development of a Bacillus subtilis expression system using the improved Pglv promoter. *Microb. Cell Fact.* **9,** 55 (2010).

381. Ming-Ming, Y., Wei-Wei, Z., Xi-Feng, Z. & Pei-Lin, C. Construction and characterization of a novel maltose inducible expression vector in Bacillus subtilis. *Biotechnol. Lett.* **28,** 1713–1718 (2006).

382. Yansura, D. G. & Hennert, D. J. Use of the Escherichia coli lac repressor and operator to control gene expression in Bacillus subtilis. *Proc. Natl. Acad. Sci. U S A* **81,** 439–443 (1984).

383. Le Grice, S. F. J. Regulated Promoter for High-Level Expression of Heterologous Genes in Bacillus subtilis. *Methods Enzymol.* **185,** (1990).

384. Esvelt, K. M. & Wang, H. H. Genome-scale engineering for systems and synthetic biology. *Mol. Syst. Biol.* **9,** 641–641 (2014).

385. Mosberg, J. A., Lajoie, M. J. & Church, G. M. Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate. *Genetics* **186,** 791–799 (2010).

386. Schweizer, H. P. Applications of the Saccharomyces cerevisiae Flp-FRT system in bacterial genetics. *J. Mol. Microbiol. Biotechnol.* **5,** 67–77 (2003).

387. Abremski, K., Wierzbicki, A., Frommer, B. & Hoess, R. H. Bacteriophage P1 Cre-loxP site-specific recombination. *J. Biol. Chem.* **261,** 391–396 (1986).

388. St-Pierre, F. *et al.* One-step cloning and chromosomal integration of DNA. *ACS Synth. Biol.* **2,** 537–541 (2013).

389. Brocchieri, L. & Karlin, S. Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res.* **33,** 3390–3400 (2005).

390. Kaur, J., Kumar, A. & Kaur, J. Strategies for optimization of heterologous protein expression in E. coli: Roadblocks and reinforcements. *Int. J. Biol. Macromol.* (2017). doi:10.1016/j.ijbiomac.2017.08.080

391. Sauvonnet, N., Poquet, I. & Pugsley, A. P. Extracellular secretion of pullulanase is unaffected by minor sequence changes but is usually prevented by adding reporter proteins to its N-or C-terminal end. *J. Bacteriol.* **177,** 5238–5246 (1995).

392. Esposito, D. & Chatterjee, D. K. Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol.* **17,** 353–358 (2006).

393. Kane, J. F. Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coil. *Curr. Opin. Biotechnol.* **6,** 494–500 (1995).
394. Wagner, S. *et al.* Consequences of Membrane Protein Overexpression in *Escherichia coli*. *Mol. Cell. Proteomics* **6,** 1527–1550 (2007).
395. Klepsch, M. M., Persson, J. O. & De Gier, J.-W. L. Consequences of the Overexpression of a Eukaryotic Membrane Protein, the Human KDEL Receptor, in Escherichia coli. *J. Mol. Biol.* **407,** 532–542 (2011).
396. Mergulhão, F. J. M. & Monteiro, G. A. Secretion capacity limitations of the Sec pathway in Escherichia coli. *J. Microbiol. Biotechnol.* **14,** 128–133 (2004).
397. Huang, H. C., Sherman, M. Y., Kandror, O. & Goldberg, A. L. The Molecular Chaperone DnaJ Is Required for the Degradation of a Soluble Abnormal Protein in Escherichia coli. *J. Biol. Chem.* **276,** 3920–3928 (2001).
398. Georgellis, D. Quinones as the Redox Signal for the Arc Two-Component System of Bacteria. *Science* **292,** 2314–2316 (2001).
399. Ellis, J. R. & Minton, A. P. Protein aggregation in crowded environments. *J. Am. Chem. Soc.* **387,** 485–497 (2006).
400. Carrió, M. M., Cubarsi, R. & Villaverde, A. Fine architecture of bacterial inclusion bodies. *FEBS Lett.* **471,** 7–11 (2000).
401. Ramón, A., Senorale-Pose, M. & Marín, M. Inclusion bodies: Not that bad... *Front. Microbiol.* **5,** (2014).
402. Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M. & Panda, A. K. Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. *Microb. Cell Fact.* **14,** 41 (2015).
403. Georgiou, G. & Valax, P. Isolating Inclusion Bodies from Bacteria. *Methods Enzymol.* **309,** 48–58 (1999).
404. Carrió, M. M. & Villaverde, A. Protein aggregation as bacterial inclusion bodies is reversible. *FEBS Lett.* **489,** 29–33 (2001).
405. Tobin, M. B., Gustafsson, C. & Huisman, G. W. Directed evolution: The 'rational' basis for 'irrational' design. *Curr. Opin. Struct. Biol.* **10,** 421–427 (2000).
406. Crameri, A., Raillard, S.-A., Bermudez, E. & Stemmer, W. P. C. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391,** 288–91 (1998).
407. Stemmer, W. P. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 10747–10751 (1994).
408. Rigouin, C., Nguyen, H. A., Schalk, A. M. & Lavie, A. Discovery of human-like L-asparaginases with potential clinical use by directed evolution. *Sci. Rep.* **7,** 10224 (2017).
409. Roberts, T. M. *et al.* Identification and Characterisation of a pH-stable GFP. *Sci. Rep.* **6,** 28166 (2016).
410. Taylor, J. L., Price, J. E. & Toney, M. D. Directed evolution of the substrate specificity of dialkylglycine decarboxylase. *Biochim. Biophys. Acta - Proteins Proteomics* **1854,** 146–155 (2015).
411. Axarli, I., Muleta, A. W., Chronopoulou, E. G., Papageorgiou, A. C. & Labrou, N. E. Directed evolution of glutathione transferases towards a selective glutathione-binding site and improved oxidative stability. *Biochim. Biophys. Acta - Gen. Subj.* **1861,** 3416–3428 (2017).
412. Chang, C. C. *et al.* Evolution of a cytokine using DNA family shuffling. *Nat. Biotechnol.* **17,** 793–797 (1999).
413. Ostermeier, M., Shim, J. H. & Benkovic, S. J. A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* **17,** 1205–1209 (1999).
414. Ness, J. E. *et al.* DNA shuffling of subgenomic sequences of subtilisin. *Nat. Biotechnol.* **17,** 893–896 (1999).
415. Voigt, C. A., Martinez, C., Wang, Z.-G., Mayo, S. L. & Arnold, F. H. Protein building blocks preserved by recombination. *Nat. Struct. Biol.* **9,** 553–558 (2002).
416. O'Maille, P. E., Bakhtina, M. & Tsai, M. D. Structure-based combinatorial protein engineering (SCOPE). *J. Mol. Biol.* **321,** 677–691 (2002).
417. Li, Y. *et al.* A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments. *Nat. Biotechnol.* **25,** 1051–1056 (2007).
418. Meyer, M. M., Hochrein, L. & Arnold, F. H. Structure-guided SCHEMA recombination of distantly related β-lactamases. *Protein Eng. Des. Sel.* **19,** 563–570 (2006).
419. Pedelacq, J.-D. *et al.* Engineering soluble proteins for structural genomics. *Nat. Biotechnol.* **20,** 927–932 (2002).
420. Sharpl, P. M. & Li, W. potential applications Nucleic Acids Research. **15,** 1281–1295 (1987).
421. Nørholm, M. H. H. *et al.* Manipulating the genetic code for membrane protein production: What have we learnt so far? *Biochim. Biophys. Acta - Biomembr.* **1818,** 1091–1096 (2012).

422. Gustafsson, C., Govindarajan, S. & Minshull, J. Codon bias and heterologous protein expression. *Trends Biotechnol.* **22,** 346–353 (2004).

423. Buhr, F. *et al.* Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. *Mol. Cell* **61,** 341–351 (2016).

424. Tsai, C. J. *et al.* Synonymous Mutations and Ribosome Stalling Can Lead to Altered Folding Pathways and Distinct Minima. *J. Mol. Biol.* **383,** 281–291 (2008).

425. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of gene expression in Escherichia coli. *Science* **324,** 255–258 (2009).

426. Goodman, D. B., Church, G. M. & Kosuri, S. Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science* **342,** 475–479 (2013).

427. Boël, G. *et al.* Codon influence on protein expression in E. coli correlates with mRNA levels. *Nature* **529,** 358–363 (2016).

428. Li, G. W., Oh, E. & Weissman, J. S. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484,** 538–541 (2012).

429. Martinez, A. *et al.* Expression of recombinant human phenylalanine hydroxylase as fusion protein in Escherichia coli circumvents proteolytic degradation by host cell proteases. Isolation and characterization of the wild-type enzyme. *Biochem. J.* **306,** 589–597 (1995).

430. Jacquet, a *et al.* Expression of a recombinant Toxoplasma gondii ROP2 fragment as a fusion protein in bacteria circumvents insolubility and proteolytic degradation. *Protein Expr. Purif.* **17,** 392–400 (1999).

431. Samuelsson, E., Moks, T., Uhlen, M. & Nilsson, B. Enhanced in vitro Refolding of Insulin-like Growth Factor I Using a Solubilizing Fusion Partner. *Biochemistry* **33,** 4207–4211 (1994).

432. Terpe, K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **60,** 523–533 (2003).

433. Drew, D. *et al.* A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. *Protein Sci.* **14,** 2011–2017 (2005).

434. Drew, D., Lerch, M., Kunji, E., Slotboom, D.-J. & de Gier, J.-W. Optimization of membrane protein overexpression and purification using GFP fusions. *Nat. Methods* **3,** 303–313 (2006).

435. Waldo, G. S., Standish, B. M., Berendzen, J. & Terwilliger, T. C. Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* **17,** 691–5 (1999).

436. Feilmeier, B. J., Iseminger, G., Schroeder, D., Webber, H. & Phillips, G. J. Green fluorescent protein functions as a reporter for protein localization in Escherichia coli. *J. Bacteriol.* **182,** 4068–76 (2000).

437. Drew, D. *et al.* Rapid topology mapping of Escherichia coli inner-membrane proteins by prediction and PhoA/GFP fusion analysis. *Proc. Natl. Acad. Sci. U S A* **99,** 2690–2695 (2002).

438. Drew, D. E., von Heijne, G., Nordlund, P. & de Gier, J. W. L. Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli. *FEBS Lett.* **507,** 220–224 (2001).

439. Daley, D. O. *et al.* Global topology analysis of the Escherichia coli inner membrane proteome. *Science* **308,** 1321–1323 (2005).

440. Toddo, S. *et al.* Application of split-green fluorescent protein for topology mapping membrane proteins in Escherichia coli. *Protein Sci.* **21,** 1571–1576 (2012).

441. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24,** 79–88 (2006).

442. Kawate, T. & Gouaux, E. Fluorescence-Detection Size-Exclusion Chromatography for Precrystallization Screening of Integral Membrane Proteins. *Structure* **14,** 673–681 (2006).

443. Parcej, D., Guntrum, R., Schmidt, S., Hinz, A. & Tampé, R. Multicolour Fluorescence-Detection Size-Exclusion Chromatography for Structural Genomics of Membrane Multiprotein Complexes. *PLoS One* **8,** e67112 (2013).

444. Skretas, G. & Georgiou, G. Genetic analysis of G protein-coupled receptor expression in Escherichia Coli: Inhibitory role of DnaJ on the membrane integration of the human central cannabinoid receptor. *Biotechnol. Bioeng.* **102,** 357–367 (2009).

445. Daley, D. O. *et al.* Global Topology Analysis of the Escherichia coli Inner Membrane Proteome. *Science* **308,** 1321–1323 (2005).

446. Skretas, G. & Georgiou, G. Simple Genetic Selection Protocol for Isolation of Overexpressed Genes That Enhance Accumulation of Membrane-Integrated Human G Protein-Coupled Receptors in Escherichia

coli. *Appl. Environ. Microbiol.* **76,** 5852–5859 (2010).

447. Natarajan, A., Haitjema, C. H., Lee, R., Boock, J. T. & DeLisa, M. P. An Engineered Survival-Selection Assay for Extracellular Protein Expression Uncovers Hypersecretory Phenotypes in Escherichia coli. *ACS Synth. Biol.* **6,** 875–883 (2017).
448. Gromek, K. A. *et al.* Improved expression and purification of sigma 1 receptor fused to maltose binding protein by alteration of linker sequence. *Protein Expr. Purif.* **89,** 203–209 (2013).
449. Vu, T. T. T. *et al.* Soluble prokaryotic expression and purification of human interferon alpha-2b using a maltose-binding protein tag. *J. Mol. Microbiol. Biotechnol.* **26,** 359–368 (2016).
450. Salema, V. & Fernández, L. Á. High yield purification of nanobodies from the periplasm of E. coli as fusions with the maltose binding protein. *Protein Expr. Purif.* **91,** 42–48 (2013).
451. Bach, H. *et al.* Escherichia coli maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies. *J. Mol. Biol.* **312,** 79–93 (2001).
452. Kapust, R. B. & Waugh, D. S. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8,** 1668–1674 (1999).
453. Lee, S. B., Park, S. K. & Kim, Y. S. Maltose binding protein-fusion enhances the bioactivity of truncated forms of pig myostatin propeptide produced in E. coli. *PLoS One* **12,** 1–16 (2017).
454. Reuten, R. *et al.* Maltose-binding protein (MBP), a secretion-enhancing tag for mammalian protein expression systems. *PLoS One* **11,** 1–15 (2016).
455. Maina, C. V. *et al.* An Escherichia coli vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. *Gene* **74,** 365–373 (1988).
456. Bedouelle, H. & Duplay, P. Production in Escherichia coli and one-step purification of bifunctional hybrid proteins which bind maltose: Export of the Klenow polymerase into the periplasmic space. *Eur. J. Biochem.* **171,** 541–549 (1988).
457. Hu, J., Qin, H., Gao, F. P. & Cross, T. A. A systematic assessment of mature MBP in membrane protein production: Overexpression, membrane targeting and purification. *Protein Expr. Purif.* **80,** 34–40 (2011).
458. Smith, D. B. & Johnson, K. S. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene* **67,** 31–40 (1988).
459. Zhao, X. *et al.* Several Affinity Tags Commonly Used in Chromatographic Purification. *J. Anal. Methods Chem.* **2013,** 1–8 (2013).
460. Dyson, M. R., Shadbolt, S. P., Vincent, K. J., Perera, R. L. & McCafferty, J. Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. *BMC Biotechnol.* **4,** 32 (2004).
461. Marblestone, J. G. Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci.* **15,** 182–189 (2006).
462. Malakhov, M. P. *et al.* SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics* **5,** 75–86 (2004).
463. Baker, R. T. Protein expression using ubiquitin fusion and cleavage. *Curr. Opin. Biotechnol.* **7,** 541–546 (1996).
464. Wintrode, P. L., Makhatadze, G. I. & Privalov, P. L. Thermodynamics of ubiquitin unfolding. *Proteins Struct. Funct. Bioinforma.* **18,** 246–253 (1994).
465. Rogov, V. V. *et al.* A Universal Expression Tag for Structural and Functional Studies of Proteins. *ChemBioChem* **13,** 959–963 (2012).
466. Turner, P., Holst, O. & Karlsson, E. N. Optimized expression of soluble cyclomaltodextrinase of thermophilic origin in Escherichia coli by using a soluble fusion-tag and by tuning of inducer concentration. *Protein Expr. Purif.* **39,** 54–60 (2005).
467. Dümmler, A., Lawrence, A.-M. & de Marco, A. Simplified screening for the detection of soluble fusion constructs expressed in E. coli using a modular set of vectors. *Microb. Cell Fact.* **4,** 34 (2005).
468. Davis, G. D., Elisee, C., Mewham, D. M. & Harrison, R. G. New fusion protein systems designed to give soluble expression in Escherichia coli. *Biotechnol. Bioeng.* **65,** 382–388 (1999).
469. Costa, S., Almeida, A., Castro, A. & Domingues, L. Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system. *Front. Microbiol.* **5,** 1–20 (2014).
470. Hammarström, Martin, Hellgren, Niklas, van den Berg, Susanne, Berglund, Helena and Härd, T. Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia

coli. *Protein Sci.* **11,** 313–321 (2002).

471. Corsini, L., Hothorn, M., Scheffzek, K., Sattler, M. & Stier, G. Thioredoxin as a fusion tag for carrier-driven crystallization. *Protein Sci.* **17,** 2070–2079 (2008).
472. Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J. & Tan, S. Comparison of affinity tags for protein purification. *Protein Expr. Purif.* **41,** 98–105 (2005).
473. Park, W. J., You, S. H., Choi, H. A., Chu, Y. J. & Kim, G. J. Over-expression of recombinant proteins with N-terminal His-tag via subcellular uneven distribution in Escherichia coli. *Acta Biochim. Biophys. Sin. (Shanghai).* **47,** 488–495 (2015).
474. Maertens, B., Spriestersbach, A., Kubicek, J. & Sch??fer, F. Strep-Tagged Protein Purification. *Methods Enzymol.* **559,** 53–69 (2015).
475. Schmidt, T. G. M., Koepke, J., Frank, R. & Skerra, A. Molecular Interaction Between the Strep-tag Affinity Peptide and its Cognate Target, Streptavidin. *J. Mol. Biol.* **255,** 753–766 (1996).
476. Schmidt, T. G. M. & Skerra, A. The random peptide library-assisted engineering of a c-terminal affinity peptide, useful for the detection and purification of a functional ig Fv fragment. *Protein Eng. Des. Sel.* **6,** 109–122 (1993).
477. Schmidt, T. G. & Skerra, A. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat. Protoc.* **2,** 1528–1535 (2007).
478. Schmidt, T. G. M. *et al.* Development of the Twin-Strep-tag and its application for purification of recombinant proteins from cell culture supernatants. *Protein Expr. Purif.* **92,** 54–61 (2013).
479. Keefe, A. D., Wilson, D. S., Seelig, B. & Szostak, J. W. One-Step Purification of Recombinant Proteins Using a Nanomolar-Affinity Streptavidin-Binding Peptide, the SBP-Tag. *Protein Expr. Purif.* **23,** 440–446 (2001).
480. Hopp, T. P. *et al.* A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification. *Bio/Technology* **6,** 1204–1210 (1988).
481. Sasaki, F. *et al.* A high-affinity monoclonal antibody against the FLAG tag useful for G-protein-coupled receptor study. *Anal. Biochem.* **425,** 157–165 (2012).
482. Einhauer, A. & Jungbauer, A. The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. *J. Biochem. Biophys. Methods* **49,** 455–465 (2001).
483. Gerace, E. & Moazed, D. *Affinity Pull-Down of Proteins Using Anti-FLAG M2 Agarose Beads*. *Methods in Enzymology* **559,** (Elsevier Inc., 2015).
484. Moon, J.-M., Kim, G.-Y. & Rhim, H. A new idea for simple and rapid monitoring of gene expression: requirement of nucleotide sequences encoding an N-terminal HA tag in the T7 promoter-driven expression in E. coli. *Biotechnol. Lett.* **34,** 1841–6 (2012).
485. Hillman, M. C. *et al.* A comprehensive system for protein purification and biochemical analysis based on antibodies to c-myc peptide. *Protein Expr. Purif.* **23,** 359–68 (2001).
486. Chen, X., Zaro, J. L. & Shen, W. C. Fusion protein linkers: Property, design and functionality. *Adv. Drug Deliv. Rev.* **65,** 1357–1369 (2013).
487. Waugh, D. S. An overview of enzymatic reagents for the removal of affinity tags. *Protein Expr. Purif.* **80,** 283–293 (2011).
488. Makrides, S. C. Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiol. Rev.* **60,** 512–538 (1996).
489. Shokri, A., Sandén, A. & Larsson, G. Cell and process design for targeting of recombinant protein into the culture medium of Escherichia coli. *Appl. Microbiol. Biotechnol.* **60,** 654–664 (2003).
490. Lee, Y. J. & Jeong, K. J. Challenges to production of antibodies in bacteria and yeast. *J. Biosci. Bioeng.* **120,** 483–490 (2015).
491. Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361,** 13–37 (2005).
492. Laursen, B. S., Sørensen, H. P., Mortensen, K. K. & Sperling-Petersen, H. U. Initiation of Protein Synthesis in Bacteria. *Microbiol. Mol. Biol. Rev.* **69,** 101–123 (2005).
493. Reeve, B., Hargest, T., Gilbert, C. & Ellis, T. Predicting translation initiation rates for designing synthetic biology. *Front. Bioeng. Biotechnol.* **2,** (2014).
494. Osterman, I. A., Evfratov, S. A., Sergiev, P. V. & Dontsova, O. A. Comparison of mRNA features affecting translation initiation and reinitiation. *Nucleic Acids Res.* **41,** 474–486 (2013).

495. Schurr, T., Nadir, E. & Margalit, H. Identification and characterization of E.coli ribosomal binding sites by free energy computation. *Nucleic Acids Res.* **21,** 4019–4023 (1993).

496. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **10,** 354–360 (2013).

497. Bonde, M. T. *et al.* Predictable tuning of protein expression in bacteria. *Nat. Methods* **13,** 233–236 (2016).

498. Marino, J., Hohl, M., Seeger, M. A., Zerbe, O. & Geertsma, E. R. Bicistronic mRNAs to Enhance Membrane Protein Overexpression. *J. Mol. Biol.* **427,** 943–954 (2015).

499. Van Etten, W. J. & Janssen, G. R. An AUG initiation codon, not codon-anticodon complementarity, is required for the translation of unleadered mRNA in Escherichia coli. *Mol. Microbiol.* **27,** 987–1001 (1998).

500. Donnell, S. M. O. & Janssen, G. R. The Initiation Codon Affects Ribosome Binding and Translational Efficiency in Escherichia coli of c I mRNA with or without the 5 ′ Untranslated Leader The Initiation Codon Affects Ribosome Binding and Translational Efficiency in Escherichia coli of cI mRN. *J. Bacteriol.* **183,** 1277–1283 (2001).

501. Gren, E. J. Recognition of messenger RNA during translational initiation in Escherichia coli. *Biochimie* **66,** 1–29 (1984).

502. Schneider, T. D., Stormo, G. D., Gold, L. & Ehrenfeucht, A. Information content of binding sites on nucleotide sequences. *J. Mol. Biol.* **188,** 415–431 (1986).

503. Clark, B. F. C. & Marcker, K. A. The role of N-formyl-methionyl-sRNA in protein biosynthesis. *J. Mol. Biol.* **17,** 394–406 (1966).

504. Egbert, R. G. & Klavins, E. Fine-tuning gene networks using simple sequence repeats. *Proc. Natl. Acad. Sci. U S A* **109,** 16817–16822 (2012).

505. Chen, H., Bjerknes, M., Kumar, R. & Jay, E. Determination of the optimal aligned spacing between the shine - dalgarno sequence and the translation initiation codon of escherichia coli m RNAs. *Nucleic Acids Res.* **22,** 4953–4957 (1994).

506. Mirzadeh, K., Toddo, S., Nørholm, M. H. H. & Daley, D. O. Codon Optimizing for Increased Membrane Protein Production: A Minimalist Approach Kiavash. in *Heterologous Expression of Membrane Proteins: Methods and Protocols, Methods in Molecular Biology* **1432,** 53–61 (2016).

507. Kochetov, A. V *et al.* AUG _ hairpin : prediction of a downstream secondary structure influencing the recognition of a translation start site. *BMC Bioinformatics* **8,** (2007).

508. Jeschek, M., Gerngross, D. & Panke, S. Rationally reduced libraries for combinatorial pathway optimization minimizing experimental effort. *Nat. Commun.* **7,** 11163 (2016).

509. Stenström, C. M., Jin, H., Major, L. L., Tate, W. P. & Isaksson, L. A. Codon bias at the 3′-side of the initiation codon is correlated with translation initiation effciency in Escherichia coli. *Gene* **263,** 273–284 (2001).

510. Looman, A. C. *et al.* Influence of the codon following the AUG initiation codon expression of a modified lacZ gene in Escherichia coli. *EMBO J.* **6,** 2489–2492 (1987).

511. Bivona, L., Zou, Z., Stutzman, N. & Sun, P. D. Influence of the second amino acid on recombinant protein expression. *Protein Expr. Purif.* **74,** 248–256 (2010).

512. Nørholm, M. H. H. *et al.* Improved production of membrane proteins in escherichia coli by selective codon substitutions. *FEBS Lett.* **587,** 2352–2358 (2013).

513. Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z. & Blüthgen, N. Efficient translation initiation dictates codon usage at gene start. *Mol. Syst. Biol.* **9,** 1–10 (2013).

514. Espah Borujeni, A. & Salis, H. M. Translation Initiation is Controlled by RNA Folding Kinetics via a Ribosome Drafting Mechanism. *J. Am. Chem. Soc.* **138,** 7016–7023 (2016).

515. Tuller, T. & Zur, H. Multiple roles of the coding sequence 5′ end in gene expression regulation. *Nucleic Acids Res.* **43,** 13–28 (2015).

516. Tuller, T. *et al.* An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. *Cell* **141,** 344–354 (2010).

517. Cheong, D.-E. *et al.* Enhancing Functional Expression of Heterologous Proteins Through Random Substitution of Genetic Codes in the 5 ′ Coding Region. *Biotechnol. Bioeng.* **112,** 822–826 (2015).

518. Salis, H. M. The ribosome binding site calculator. *Methods Enzymol.* **498,** 19–42 (2011).

519. Espah Borujeni, A., Channarasappa, A. S. & Salis, H. M. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic*

Acids Res. **42,** 2646–2659 (2014).

520. Reis, A. C. & Salis, H. An automated model test system for systematic development and improvement of gene expression models. *bioRxiv* (2017). doi:10.1101/193367
521. Farasat, I. *et al.* Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Mol. Syst. Biol.* **10,** 731–731 (2014).
522. Ng, C. Y., Farasat, I., Maranas, C. D. & Salis, H. M. Rational design of a synthetic Entner-Doudoroff pathway for improved and controllable NADPH regeneration. *Metab. Eng.* **29,** 86–96 (2015).
523. Li, T. *et al.* Semirational Approach for Ultrahigh Poly(3-hydroxybutyrate) Accumulation in Escherichia coli by Combining One-Step Library Construction and High-Throughput Screening. *ACS Synth. Biol.* **5,** 1308–1317 (2016).
524. Na, D. & Lee, D. RBSDesigner: Software for designing synthetic ribosome binding sites that yields a desired level of protein expression. *Bioinformatics* **26,** 2633–2634 (2010).
525. Seo, S. W. *et al.* Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. *Metab. Eng.* **15,** 67–74 (2013).
526. Ikemura, T. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the E. coli translational system. *J. Mol. Biol.* **151,** 389–409 (1981).
527. del Tito, B. J. *et al.* Effects of a minor isoleucyl tRNA on heterologous protein translation in Escherichia coli . Effects of a Minor Isoleucyl tRNA on Heterologous Protein Translation in Escherichia coli. *J. Bacteriol.* **177,** 7086–91 (1995).
528. Kim, R. *et al.* Overexpression of archaeal proteins in Escherichia coli. *Biotechnol. Lett.* **20,** 207–210 (1998).
529. Baca, A. M. & Hol, W. G. J. Overcoming codon bias: A method for high-level overexpression of Plasmodium and other AT-rich parasite genes in Escherichia coli. *Int. J. Parasitol.* **30,** 113–118 (2000).
530. Tegel, H., Tourle, S., Ottosson, J. & Persson, A. Increased levels of recombinant human proteins with the Escherichia coli strain Rosetta(DE3). *Protein Expr. Purif.* **69,** 159–167 (2010).
531. Søgaard, K. M. & Nørholm, M. H. H. Side effects of extra tRNA supplied in a typical bacterial protein production scenario. *Protein Sci.* **25,** 2102–2108 (2016).
532. de Marco, A., Vigh, L., Diamant, S. & Goloubinoff, P. Native folding of aggregation-prone recombinant proteins in Escherichia coli by osmolytes, plasmid- or benzyl alcohol–overexpressed molecular chaperones. *Cell Stress Chaperones* **10,** 329 (2005).
533. Voulgaridou, G. P., Mantso, T., Chlichlia, K., Panayiotidis, M. I. & Pappa, A. Efficient E. coli Expression Strategies for Production of Soluble Human Crystallin ALDH3A1. *PLoS One* **8,** e56582 (2013).
534. Chen, Y., Song, J., Sui, S. F. & Wang, D. N. DnaK and DnaJ facilitated the folding process and reduced inclusion body formation of magnesium transporter CorA overexpressed in Escherichia coli. *Protein Expr. Purif.* **32,** 221–231 (2003).
535. Link, A. J., Skretas, G., Strauch, E., Chari, N. S. & Georgiou, G. Efficient production of membrane-integrated and detergent-soluble G protein-coupled receptors in Escherichia coli. *Protein Sci.* **17,** 1857–1863 (2008).
536. Nannenga, B. L. & Baneyx, F. Reprogramming chaperone pathways to improve membrane protein expression in Escherichia coli. *Protein Sci.* **20,** 1411–1420 (2011).
537. Lamppa, J. W., Tanyos, S. A. & Griswold, K. E. Engineering Escherichia coli for soluble expression and single step purification of active human lysozyme. *J. Biotechnol.* **164,** 1–8 (2013).
538. Sonoda, H., Kumada, Y., Katsuda, T. & Yamaji, H. Effects of cytoplasmic and periplasmic chaperones on secretory production of single-chain Fv antibody in Escherichia coli. *J. Biosci. Bioeng.* **111,** 465–470 (2011).
539. Wang, R. *et al.* Engineering production of functional scFv antibody in E. coli by co-expressing the molecule chaperone Skp. *Front. Cell. Infect. Microbiol.* **3,** 1–9 (2013).
540. Zhang, Y. *et al.* Protein Expression and Purification Enhanced soluble production of cholera toxin B subunit in Escherichia coli by co-expression of SKP chaperones. *Protein Expr. Purif.* **138,** 1–6 (2017).
541. Kurokawa, Y., Yanagi, H. & Yura, T. Overexpression of protein disulfide isomerase DsbC stabilizes multiple-disulfide-bonded recombinant protein produced and transported to the periplasm in Escherichia coli. *Appl. Environ. Microbiol.* **66,** 3960–3965 (2000).
542. Lee, Y. J., Lee, D. H. & Jeong, K. J. Enhanced production of human full-length immunoglobulin G1 in the periplasm of Escherichia coli. *Appl. Microbiol. Biotechnol.* **98,** 1237–1246 (2014).

543. Maskos, K., Huber-Wunderlich, M. & Glockshuber, R. DsbA and DsbC-catalyzed oxidative folding of proteins with complex disulfide bridge patterns in vitro and in vivo. *J. Mol. Biol.* **325,** 495–513 (2003).

544. Makino, T., Skretas, G., Kang, T. H. & Georgiou, G. Comprehensive engineering of Escherichia coli for enhanced expression of IgG antibodies. *Metab. Eng.* **13,** 241–251 (2011).

545. Schlapschy, M., Grimm, S. & Skerra, A. A system for concomitant overexpression of four periplasmic folding catalysts to improve secretory protein production in Escherichia coli. *Protein Eng. Des. Sel.* **19,** 385–390 (2006).

546. Martínez-Alonso, M., García-Fruitós, E., Ferrer-Miralles, N., Rinas, U. & Villaverde, A. Side effects of chaperone gene co-expression in recombinant protein production. *Microb. Cell Fact.* **9,** 64 (2010).

547. Lindholm, A., Ellmén, U., Tolonen-Martikainen, M. & Palva, A. Heterologous protein secretion in Lactococcus lactis is enhanced by the Bacillus subtilis chaperone-like protein PrsA. *Appl. Microbiol. Biotechnol.* **73,** 904–914 (2006).

548. Drouault, S. *et al.* The Peptidyl-Prolyl Isomerase Motif Is Lacking in PmpA, the PrsA-Like Protein Involved in the Secretion Machinery of Lactococcus lactis Sophie. *Appl. Environ. Microbiol.* **68,** 3932–3942 (2002).

549. Chen, J. *et al.* Enhanced extracellular production of α-amylase in Bacillus subtilis by optimization of regulatory elements and over-expression of PrsA lipoprotein. *Biotechnol. Lett.* **37,** 899–906 (2015).

550. Wagner, S. *et al.* Tuning Escherichia coli for membrane protein overexpression. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 14371–14376 (2008).

551. Miroux, B. & Walker, J. E. Over-production of Proteins inEscherichia coli: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. *J. Mol. Biol.* **260,** 289–298 (1996).

552. Baumgarten, T. *et al.* Isolation and characterization of the E. coli membrane protein production strain Mutant56(DE3). *Sci. Rep.* **7,** 45089 (2017).

553. Kim, S. K., Lee, D., Kim, O. C., Kim, J. F. & Yoon, S. H. Tunable control of an Escherichia coli expression system for the overproduction of membrane proteins by titrated expression of a mutant lac repressor. *ACS Synth. Biol.* **6,** 1766–1773 (2017).

554. Prinz, W. A., Åslund, F., Holmgren, A. & Beckwith, J. The Role of the Thioredoxin and Glutaredoxin Pathways in Reducing Protein Disulfide Bonds in the Escherichia coli Cytoplasm. *J. Biol. Chem.* **272,** 15661–15667 (1997).

555. Derman, A., Prinz, W. A., Belin, D. & Beckwith, J. Mutations That Allow Disulfide Bond Formation in the Cytoplasm of Escherichia coli. *Science* **262,** 1744–1747 (1993).

556. Lobstein, J. *et al.* SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Fact.* **15,** 124 (2012).

557. Robinson, M.-P. *et al.* Efficient expression of full-length antibodies in the cytoplasm of engineered bacteria. *Nat. Commun.* **6,** 8072 (2015).

558. Kong, B. & Guo, G. L. Soluble expression of disulfide bond containing proteins FGF15 and FGF19 in the cytoplasm of Escherichia coli. *PLoS One* **9,** 1–8 (2014).

559. Mizrachi, D. *et al.* A water-soluble DsbB variant that catalyzes disulfide-bond formation in vivo. *Nat. Chem. Biol.* **13,** 1022–1028 (2017).

560. Matos, C. F. R. O. *et al.* Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in Escherichia coli CyDisCo strains. *Biotechnol. Prog.* **30,** 281–290 (2014).

561. Mesecke, N. *et al.* A disulfide relay system in the intermembrane space of mitochondria that mediates protein import. *Cell* **121,** 1059–1069 (2005).

562. Lee, J. E., Hofhaus, G. & Lisowsky, T. Erv1p from Saccharomyces cerevisiae is a FAD-linked sulfhydryl oxidase. *FEBS Lett.* **477,** 62–66 (2000).

563. Vitu, E., Bentzur, M., Lisowsky, T., Kaiser, C. A. & Fass, D. Gain of Function in an ERV/ALR Sulfhydryl Oxidase by Molecular Engineering of the Shuttle Disulfide. *J. Mol. Biol.* **362,** 89–101 (2006).

564. Hatahet, F., Nguyen, V. D., Salo, K. E. H. & Ruddock, L. W. Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of E. coli. *Microb. Cell Fact.* **9,** 67 (2010).

565. Nguyen, V. *et al.* Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of E.coli. *Microb. Cell Fact.* **10,** 1 (2011).

566. Wacker, M. N-Linked Glycosylation in Campylobacter jejuni and Its Functional Transfer into E. coli. *Science* **298,** 1790–1793 (2002).

567. Wacker, M. *et al.* Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proc. Natl. Acad. Sci.* **103,** 7088–7093 (2006).

568. Lizak, C., Fan, Y.-Y., Weber, T. C. & Aebi, M. *N* -Linked Glycosylation of Antibody Fragments in *Escherichia coli*. *Bioconjug. Chem.* **22,** 488–496 (2011).

569. Valderrama-Rincon, J. D. *et al.* An engineered eukaryotic protein glycosylation pathway in Escherichia coli. *Nat. Chem. Biol.* **8,** 434–436 (2012).

570. Jaffé, S. R. P., Strutton, B., Levarski, Z., Pandhal, J. & Wright, P. C. Escherichia coli as a glycoprotein production host: Recent developments and challenges. *Curr. Opin. Biotechnol.* **30,** 205–210 (2014).

571. Nothaft, H. & Szymanski, C. M. Protein glycosylation in bacteria: sweeter than ever. *Nat. Rev. Microbiol.* **8,** 765–778 (2010).

572. Dietrich, J. A., McKee, A. E. & Keasling, J. D. *High-Throughput Metabolic Engineering: Advances in Small-Molecule Screening and Selection*. *Annual Review of Biochemistry* **79,** (2010).

573. Aksoy, S., Squires, C. L. & Squires, C. Translational Coupling of the trpB and trpA Genes in the Escherichia coli Tryptophan Operon. *J. Bacteriol.* **157,** 363–367 (1984).

574. Mendez-Perez, D., Gunasekaran, S., Orler, V. J. & Pfleger, B. F. A translation-coupling DNA cassette for monitoring protein translation in Escherichia coli. *Metab. Eng.* **14,** 298–305 (2012).

575. Gul, N., Linares, D. M., Ho, F. Y. & Poolman, B. Evolved escherichia coli strains for amplified, functional expression of membrane proteins. *J. Mol. Biol.* **426,** 136–149 (2014).

576. Michener, J. K. & Smolke, C. D. High-throughput enzyme evolution in Saccharomyces cerevisiae using a synthetic RNA switch. *Metab. Eng.* **14,** 306–316 (2012).

577. Binder, S. *et al.* A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level. *Genome Biol.* **13,** R40 (2012).

578. Yang, G. & Withers, S. G. Ultrahigh-throughput FACS-based screening for directed enzyme evolution. *ChemBioChem* **10,** 2704–2715 (2009).

579. Tracy, B. P., Gaida, S. M. & Papoutsakis, E. T. Flow cytometry for bacteria: Enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes. *Curr. Opin. Biotechnol.* **21,** 85–99 (2010).

580. Sjostrom, S. L. *et al.* High-throughput screening for industrial enzyme production hosts by droplet microfluidics. *Lab Chip* **14,** 806–813 (2014).

581. van Sint Fiet, S., van Beilen, J. B. & Witholt, B. Selection of biocatalysts for chemical synthesis. *Proc. Natl. Acad. Sci.* **103,** 1693–1698 (2006).

582. Schallmey, M., Frunzke, J., Eggeling, L. & Marienhagen, J. Looking for the pick of the bunch: High-throughput screening of producing microorganisms with biosensors. *Curr. Opin. Biotechnol.* **26,** 148–154 (2014).

583. Suess, B., Fink, B., Berens, C., Stentz, R. & Hillen, W. A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res.* **32,** 1610–1614 (2004).

584. Weigand, J. E. & Suess, B. Tetracycline aptamer-controlled regulation of pre-mRNA splicing in yeast. *Nucleic Acids Res.* **35,** 4179–4185 (2007).

585. Buskirk, A. R., Landrigan, A. & Liu, D. R. Engineering a Ligand-Dependent RNA Transcriptional Activator. *Chem. Biol.* **11,** 1157–1163 (2004).

586. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346,** 818–822 (1990).

587. Santos, C. N. S. & Stephanopoulos, G. Melanin-based high-throughput screen for L-tyrosine production in Escherichia coli. *Appl. Environ. Microbiol.* **74,** 1190–1197 (2008).

Publications

Paper 1

TARSyn: Tuneable antibiotic resistance devices enabling bacterial synthetic evolution and protein production

Paper 2

Selection of highly expressed gene variants in *Escherichia coli* using translationally-coupled antibiotic selection markers

Paper 3

A synthetic biology approach for selection of highly expressed gene variants in Gram-positive bacteria

Paper 4

SEVA Linkers: A Versatile and Automatable DNA Backbone Exchange Standard for Synthetic Biology

Paper 1

TARSyn: Tuneable antibiotic resistance devices enabling bacterial synthetic evolution and protein production

Rennig M.[1], Martinez V.[1], Mirzadeh K.[2,3], Dunas F.[4], Röjsäter B.[4], Daley D.O.[2,3], Nørholm M.H.H.[1,3]

[1]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark; [2]Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, Sweden; [3]CloneOpt AB, Upplands Väsby, Sweden; [4]Affibody AB, Solna, Sweden

ACS Synthetic Biology 2017
DOI: 10.1021/acssynbio.7b00200

Abstract

Evolution can be harnessed to optimize synthetic biology designs. A prominent example is recombinant protein production - a dominating theme in biotechnology for more than three decades. Typically, a protein coding sequence (cds) is recombined with genetic elements, such as promoters, ribosome binding sites and terminators, which control expression in a cell factory. A major bottleneck during production is translational initiation. Previously we identified more effective translation initiation regions (TIRs) by creating sequence libraries and then selecting for a TIR that drives high-level expression – an example of synthetic evolution. However, manual screening limits the ability to assay expression levels of all putative sequences in the libraries. Here we have solved this bottleneck by designing a collection of translational coupling devices based on a RNA secondary structure. Exchange of different sequence elements in this device allows for different coupling efficiencies therefore giving the devices a tuneable nature. Sandwiching these devices between the cds and a antibiotic selection marker that function over a broad dynamic range of antibiotic concentrations, adds to the tunability and allows expression levels in large clone libraries to be probed using a simple cell survival assay on the respective antibiotic. The power of the approach is demonstrated by substantially increasing production of two commercially interesting proteins, a Nanobody® and an Affibody®. The method is a simple and inexpensive alternative to advanced screening techniques that can be carried out in any laboratory.

Keywords: *protein production, antibiotic resistance, translational coupling, selection system, synthetic evolution, translation initiation region*

Introduction

Proteins are essential for many biomedical and biotechnological applications. Most proteins are recombinantly produced in a heterologous organism, as their natural sources are impractical for high production yields[1,2,3]. In general, high yields are desirable, as this facilitates downstream applications and reduces the production costs. Furthermore, the use of recombinant sources allows facile engineering of protein modifications for increasing the activity, solubility[4,5,6] or enabling purification[7,8].

Translation initiation is often the rate-limiting step when it comes to protein production in bacteria[9,10,11]. The efficiency of translation initiation is determined by the nucleotide sequence of the

translation initiation region (TIR), which consists of the Shine-Dalgarno (SD) sequence, the region upstream of the SD which contains the translational standby site[12], and a region that stretches from the SD to the 5th codon of the cds[13]. In clones engineered for protein production, TIR is typically a random sequence formed at the junction of the vector and the first five codons of the coding sequence[14]. Such a randomly formed TIR has not co-evolved with the ribosomes of the host cell and often does not function efficiently for maximum production. Nucleotide changes in this region can affect protein production levels by as much as 1000-fold[14,15] as they affect mRNA secondary structure and binding of the ribosome – a relaxed mRNA structure being more favourable for efficient translation initiation[16,17,18,19,20]. Computational tools that use thermodynamic principles to predict an optimized TIR are available[21,22,23,24,25,26], but in the end experimental validation of the TIRs is still required. An alternative approach is to experimentally construct a TIR library and identify an optimal TIR by screening expression levels. However, the bottle-neck in this approach is the experimental evaluation of expression levels in large libraries.

To achieve optimal production yields in a recombinant setting, we have explored a synthetic evolution approach, whereby we carry out an in vitro randomisation of the TIR and then experimentally select a nucleotide sequence that is compatible with high protein production. In a recent study, partial sequence variation was introduced in the TIRs for two difficult-to-express membrane proteins that were fused to fluorescent reporters and the plasmid libraries were screened by fluorescence activated cell sorting (FACS). In libraries containing approximately 30,000 TIRs we observed up to 1000-fold variation in expression levels[14]. Still, identification of the highest performing constructs was time consuming and required access to advanced screening equipment.

An attractive alternative to screening individual clones in a library is to couple expression of the cds to a growth-based selection system. The ideal selection system is independent of the cds, and has a large dynamic screening range so that clones producing at industrially relevant titres can be identified. Previous work has used antibiotic selection markers that are either directly fused, or translationally-coupled to the gene of interest[27,28,29,30]. Translational coupling is a naturally occurring phenomenon in bacteria and is known to involve translating ribosomes that are able to resolve secondary structures that sequester the SD sequence of downstream genes[31] (Fig. 1a). However, to our knowledge, no previous study has systematically addressed the dynamic range of either the antibiotic resistance markers or the translational coupling devices – i.e. to what extend

protein production levels correlate with the host resistance to the corresponding antibiotics. Here, we explore the concentration range of six antibiotic resistance markers and we design translational coupling devices that increase this range for one of the markers.

Results and Discussion

Probing the tunability of common antibiotic resistance markers

Our initial aim was to determine which antibiotic resistance markers were suitable to use as reliable reporters for protein production. We engineered a series of clones in which a rhamnose-inducible promoter (P_{rhaBAD}) controlled the expression of a gene encoding a fluorescent protein (Dasher-GFP), a translational coupling device, and one of six commonly used antibiotic resistance markers – conferring resistance to ampicillin/carbenicillin, chloramphenicol, spectinomycin, kanamycin, gentamicin or tetracycline, respectively (Fig. 1b). The constructs were transformed into *Escherichia coli* MG1655 and expression was induced with different concentrations of rhamnose. The different resistance genes were analysed with regards to the correlation between protein production (as determined by fluorescence) and resistance to the chosen antibiotic (as determined by the Minimal Inhibitory Concentration, MIC) using a fluorescence microplate reader as illustrated in Fig. 1b. Fig. 1c shows that all constructs expressed Dasher-GFP in a rhamnose-dependent manner. All of the resistance markers, except for gentamicin, showed a linear correlation between the MIC and fluorescence ($R^2>0.9$) (Fig. 1d-i). However, the ampicillin resistance marker (β-lactamase) displayed the largest dynamic range (Fig. 1c, right) and it was therefore selected for further characterization. Viable constructs based on the tetracycline resistance marker were only recovered at very low concentrations of tetracycline. For the chloramphenicol resistance marker, no linear correlation was found above an inducer concentration of 10 mM rhamnose, indicating that the cat gene is only tuneable to a certain extent.

Exploration of an alternative translational coupling architecture

With the coupling device that we used in the previous experiment, the MIC of ampicillin began to plateau at 10mM rhamnose (Fig. 1c, right). Moreover, very high antibiotic concentrations were required for selections when expression was induced with more than 10 mM rhamnose, which could make downstream applications complicated and expensive. This limits the use of the coupling to antibiotic resistance markers when expression levels are high. To extend the selectable range of β-lactamase we explored alternative architectures of the translational coupling device.

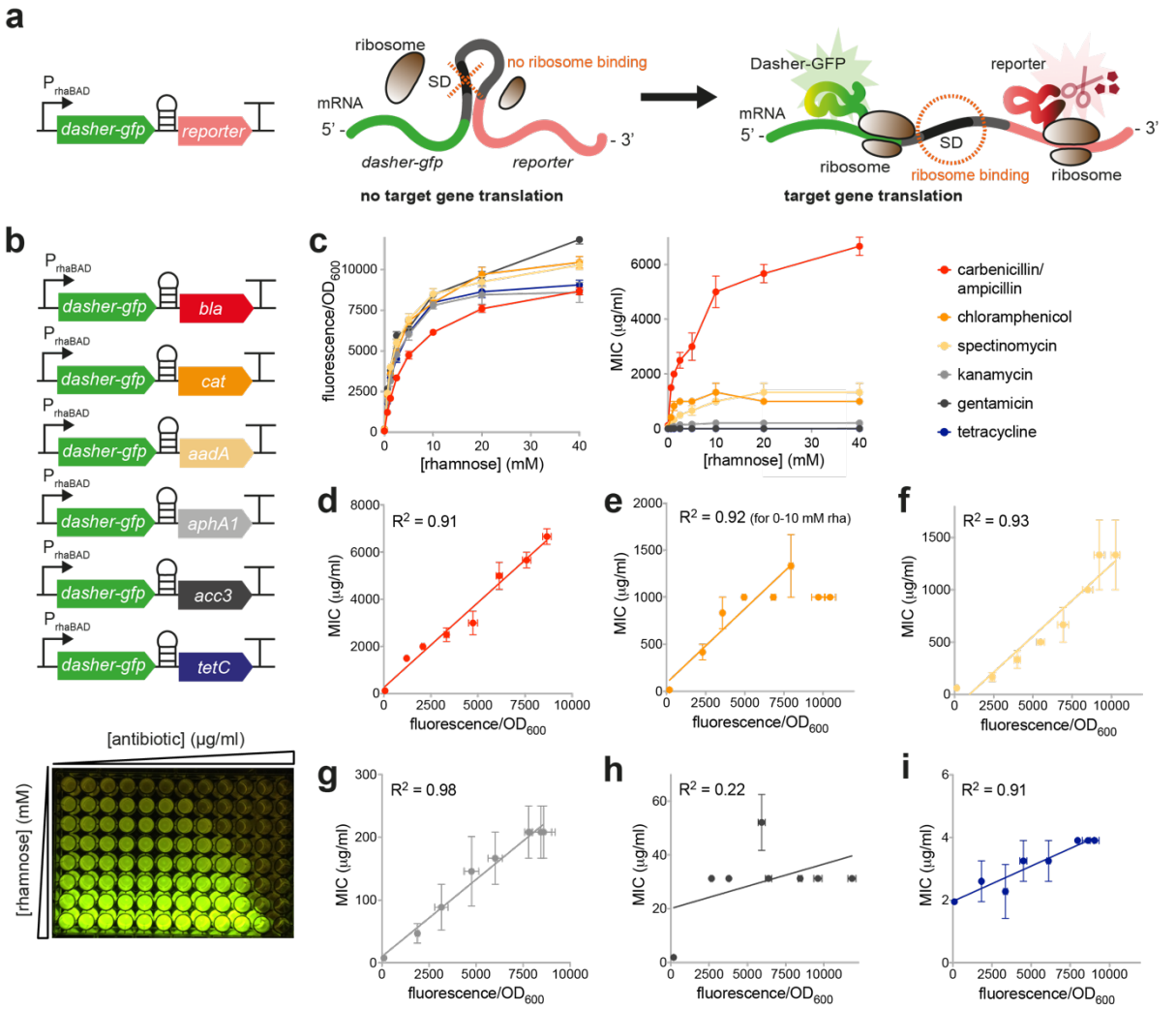


Figure 1. Comparison of different antibiotic resistance genes as translationally coupled expression reporters. (a) A cartoon demonstrating the principle of translational coupling. A gene of interest (here *dasher-gfp*, green) is cloned upstream of a translational coupling device and a reporter gene (light red). If the gene of interest is not translated, the Shine-Dalgarno (SD) sequence of the reporter gene will be masked by secondary mRNA structure and no translation will take place. If the gene of interest is translated, the secondary structure will be unfolded by the ribosome's helicase activity and the SD sequence will be accessible for a ribosome to bind and translate the reporter gene. (b) The six different expression constructs used in this study. In these constructs the gene encoding the green fluorescent protein Dasher-GFP (green) was translationally coupled with different antibiotic resistance genes and expression controlled by a rhamnose inducible promoter ($Prha_{BAD}$). Thus expression could be tuned with different rhamnose concentrations and followed during growth by simple whole-cell fluorescence in a microplate reader (illustrative example in bottom panel). In minimum inhibitory concentration (MIC) assays different rhamnose concentrations were added to simulate different expression levels and growth at different concentrations of antibiotics was tested. The investigated antibiotic resistance genes confer resistance to ampicillin *(bla*, red), chloramphenicol (*cat*, orange), spectinomycin (*aadA*, yellow), kanamycin (*aphA1*, light grey), gentamicin (*acc3*, dark grey) and tetracycline (*tetC*, blue). (c) To assure that induction was similar in all constructs, fluorescence was plotted against inducer concentration (rhamnose) (left panel). To analyse the dynamic range of each antibiotic resistance gene the obtained MIC values (colour coded for each gene as indicated above) were plotted against inducer concentration (rhamnose) (right panel). Finally, the correlation between protein production, determined as fluorescence normalized for cell density, with MIC values was determined for ampicillin (d), chloramphenicol (e), spectinomycin (f), kanamycin (g), gentamicin (h) and tetracycline (i).

Our goal here was to design a device with a weaker output from β-lactamase, so that less ampicillin could be used in the selections (Fig. 2a). Since coupling efficiency must be dependent on proximity to the upstream gene we rationally designed hairpin structures where the stop codon of the upstream gene was located at the base of a hairpin that contained the TIR of the downstream gene (Fig. 2b). This way, the two coding sequences were in close vicinity, but the hairpin structure was not part of the cds of the upstream gene. The complementary part of the stem region encodes for the second amino acid of the downstream gene (Fig. 2b). We chose the UGA stop codon, as the ampicillin and kanamycin resistance gene comprise a serine as the second amino acid, encoded by UCA, complementary to the UGA stop codon. The addition of serine had no effect on the other resistance markers tested. The three nucleotides following the stop codon in the stem region are determined by the choice of start codon in the complementary downstream region and two of these together with the next seven nucleotides comprise the Shine-Dalgarno (SD) sequence for the reporter gene. We chose a SD sequence consisting of the nine nucleotides AUAGGAGGU. The stem region was designed to contain 11 nucleotide pairs in total.

To reduce the translation efficiency of β-lactamase whilst preserving the overall hairpin structure, we designed two alternative hairpins. In the first we exchanged the AUG start codon with the less efficiently translated AUU start codon (Fig. 2b, indicated with an asterisk). Coupling efficiency is not only reduced by using the weaker start codon but potentially also by stronger termination of translation of the target gene, due to a simultaneous change of the base following the stop codon most likely causing rapid selection and interaction of the release factor with the stop codon avoiding frame shifts, as previously shown[32]. To further reduce coupling efficiency, we additionally modified the Shine-Dalgarno sequence so that it was less complementary to the 16S rRNA. We changed the very strong sequence to the very weak AUUCCUGGU sequence for the most effective reduction of coupling efficiency (Fig. 2b, indicated with an asterisk).

The hairpins were designed with a very small loop region connecting the two halves of the stem. The loop region consists of only three nucleotides where two nucleotides are defined by the chosen SD sequence. The small loop region was chosen so that the spacing between the SD sequence and the start codon of the reporter gene is close to the optimal five nucleotides[33]. Due to the experimentally determined good correlation coefficients between 0.87 and 0.96 that we achieved with the small loop region, we preferred to not elongate the loop region. The hairpin folding

propensity was tested in silico with the mfold web server[34]. The ΔG of the hairpins varied between -14.5 (strong coupling), -12.4 (weak coupling 1) and -12.3 (weak coupling 2).

Determination of the MICs at different expression levels, confirmed that these designs reduced the levels of antibiotic resistance, while preserving high correlation co-efficiency between expression of the upstream gene and the MIC (Fig 2c-d). For example, when using the coupling device with the AUU start codon, the MIC value was reduced from 3 to 0.2 mg/ml ampicillin with 5 mM rhamnose and from 6.5 to 0.5 mg/ml with 40 mM rhamnose. When using the weak coupling device with the AUU start codon and the suboptimal Shine-Dalgarno sequence, the MIC value was lowered to 0.1 mg/ml ampicillin at 5 mM rhamnose induction and to 0.25 mg/ml ampicillin at 40 mM rhamnose. The detected fluorescence levels stayed similar in all three coupling set-ups indicating that the expression of the upstream gene is unaffected by the use of different coupling devices (Fig. 2d). Thus these experiments demonstrate that translational coupling devices can be used to tune the range of an antibiotic resistance marker.

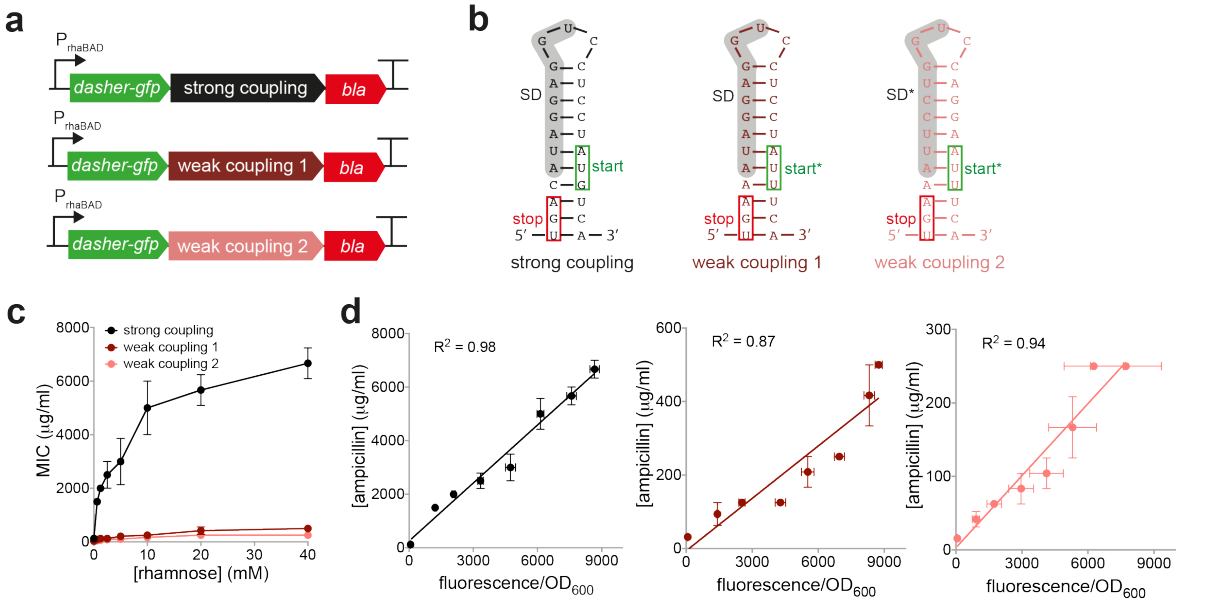


Figure 2. Modifying the translational coupling efficiency by rational sequence changes in the predicted RNA hairpin structure. (a) DNA sequences with strong and weak coupling efficiencies between the upstream gene (here dasher-gfp, green colour) and the downstream selection marker (here the ampicillin resistance gene, red colour) were designed to enable selection under different expression regimes. (b) Sequences and predicted structures of the strong and weak translational coupling devices. Important features are boxed: stop codon of the upstream gene (red), start codons of the downstream genes (AUG or AUU, green) and Shine-Dalgarno (SD) sequences (AUAGGAGGU or AUUCCUGGA, grey). Differences are indicated with an asterisk. (c) In MIC assays, different inducer concentrations were added to simulate different expression levels of dasherGFP and resistance towards different concentrations of ampicillin were tested for the strong (black) and the two weak coupling devices (dark and light red). (d) Correlation between protein production, determined as fluorescence normalized for cell density, with MIC values was determined for the strong (black) and the two weak translational coupling devices (dark and light red).

Antibiotic-resistance-based selection of high expressers from TIR variation libraries

We next explored the combination of a translational coupling devices and β-lactamase in the selection of optimal TIRs from sequence-randomized libraries. We focused on two well-studied genes, *araH* and *narK*, that are highly sensitive to TIR variation and whose expression can be estimated by fusions to GFP[14,35]. The regions encoding NarK-GFP and AraH-GFP were first fused to a translational coupling device and β-lactamase (Fig. 3a). Previously characterized NarK-GFP and AraH-GFP expression variants were used to validate the use of β-lactamase as an indicator of protein production levels (Supplementary Fig. S1a). We then created TIR libraries by randomizing the six nucleotides upstream from the start codon and sampling all synonymous codons in the first two positions after the start codon as described previously[14]. These libraries were designed to contain circa 30,000 TIR variants. The libraries were transformed into *E. coli* BL21(DE3) pLysS, protein production induced with IPTG and the cultures plated on agar plates with three different concentrations of ampicillin (Fig. 3b). This was repeated in parallel cultures with different ampicillin concentrations and each culture was analysed by flow cytometry (Fig. 3b). The cell library populations exhibited higher average fluorescence intensities as a result of growth in higher concentrations of ampicillin (Fig. 3C). When plating the libraries directly on LB agar plates containing different concentrations of ampicillin, we observed a decrease in the number of colonies but an increase in average colony fluorescence with rising antibiotic concentrations (shown for AraH-GFP in Fig. 3d). Individual colonies were picked from these plates and assayed for expression levels assessed by fluorescence per cell density. *araH-gfp* and *nark-gfp* variants selected on the highest antibiotic concentration showed to also be the highest expressing variants with 25-fold and 10-fold increased expression, respectively (Supplementary Fig. S1b). This demonstrates that synthetic evolution of the TIR is resulting in populations enriched in high expressing clones and that our tuneable translation coupling devices enable identification of those with substantially increased protein production capacities.

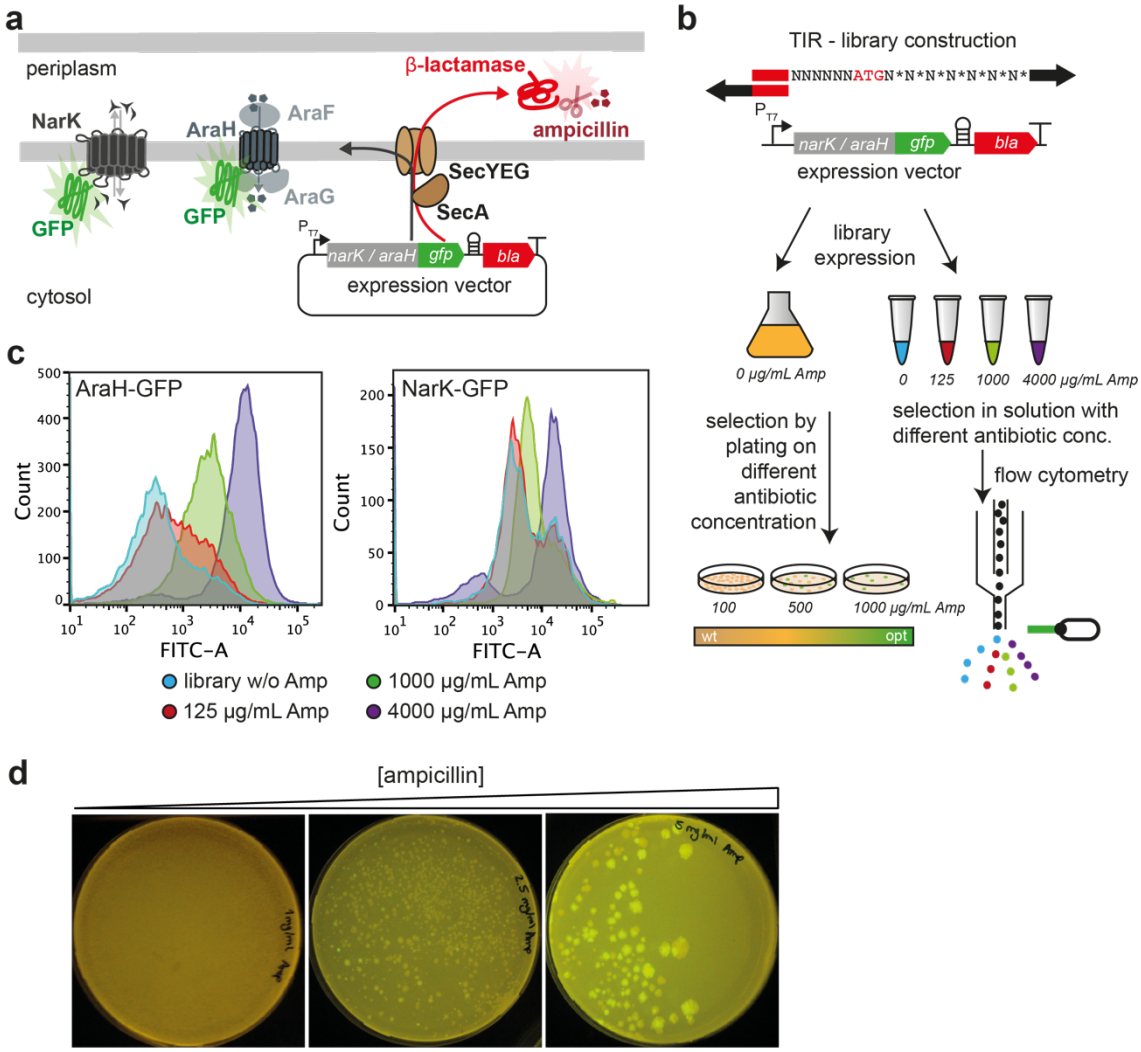


Figure 3. Enriching populations of highly expressing gene variants from a TIR optimization library by screening for resistance to ampicillin. (a) Expression of the two *E. coli* membrane transporters NarK and AraH is highly sensitive to TIR variation and can be followed by direct fusions to GFP. In this study these fusions were translationally coupled to β-lactamase. All three proteins use the SecYEG translocon for targeting to the inner membrane or the periplasm. (b) NarK-GFP and AraH-GFP TIR libraries were constructed by PCR, transformed into BL21(DE3) pLysS and grown with different ampicillin concentrations either directly in liquid culture or on agar plates. Expression levels of the different synthetically evolved populations were assayed by flow cytometry (c) or by exposing agar plates to UV light (d). Selection on plates is here demonstrated for AraH-GFP.

Optimization of Nanobody® production by synthetic evolution of the TIR

To demonstrate the effectiveness of the tuneable antibiotic resistance cassettes we used them to identify optimal TIRs for commercially interesting proteins. Initially we focussed on a camelide-derived single chain antibody (also known as a Nanobody®, Fig. 4a) that binds to GFP and enhances its fluorescence[36,37]. To identify a TIR that drove high-level expression of the Nanobody® we fused the cds to strong and weak coupling devices and the β-lactamase gene thereby creating two parallel constructs (Fig. 4b). As we had no estimate of the initial expression level, we hypothesized that the construction and selection of two parallel constructs would increase the odds

for identifying optimized TIRs with minimal additional effort. TIR libraries, each containing circa 30,000 clones, were generated and transformed into BL21(DE3) pLysS, induced with IPTG and plated onto LB agar containing increasing concentrations of ampicillin (Fig. 4c). We observed that, regardless of the coupling device, the TIR library produced viable colonies at higher ampicillin concentrations than the original TIR (Fig. 4d, e). And as expected, the weak coupling device was more sensitive to ampicillin. Plasmids were recovered from five different colonies selected on 1.5 mg/mL and 4 mg/mL of ampicillin for the weak and the strong coupling device, respectively, and two unique TIRs were identified – one from each experiment. Western blotting confirmed that the increase in antibiotic resistance of the synthetically evolved constructs was paralleled by an increase in the production of the Nanobody® (Fig. 4f). Importantly, for both of the synthetically optimized clones, the activity of the Nanobody® was preserved as judged by its ability to bind GFP (Fig. 4g). Taken together the production of the Nanobody® could be improved 10-fold and 3-fold using weak and strong coupling devices, respectively. This exemplifies the value of different translational coupling devices to tune the resistance of the antibiotic resistance reporter. The yield of purified Nanobody®/L of culture was estimated around 986 mg and 86 mg using the weak and strong coupling devices, respectively, judged by densiometric analysis in accordance with a standard curve of purified protein of known concentration (Supplementary Fig. S3a, b). Sequences of the selected TIR variants are shown in Table S4.

Optimization of Affibody® production by synthetic evolution of the TIR

We also explored the effectiveness of the tuneable antibiotic resistance cassettes for identifying optimal TIRs for an Affibody® molecule. Affibody® molecules are small (6.5-kDa) affinity proteins based on a three-helix bundle domain framework[38]. We focussed on an albumin binding domain (ABD)-fused dimeric Affibody® molecule (Fig. 5a). This molecule binds to the human epidermal growth factor receptor 3 (HER3), which is implicated in a number of different cancers[39]. In the experiment we translationally coupled the cds to the β-lactamase gene and then created a TIR library (Fig. 5b). The TIR library contained circa 30,000 clones, which were transformed into BL21(DE3) pLysS, induced with IPTG and plated onto LB agar containing increasing concentrations of ampicillin. We observed that the TIR library produced viable colonies at higher ampicillin concentrations than the original TIR (Fig. 5c).

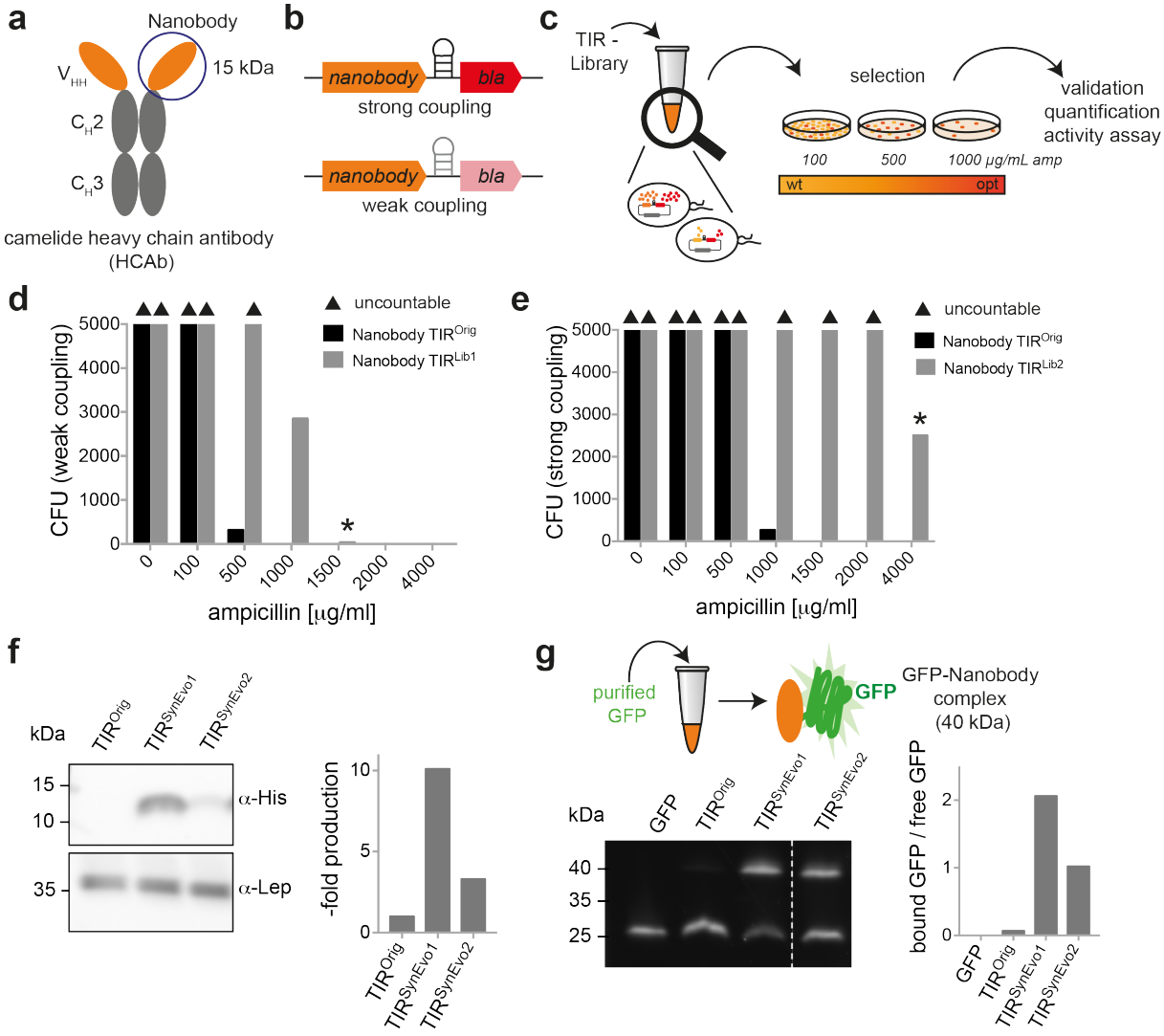


Figure 4. Optimized production of a single chain antibody by synthetic evolution. (a) Nanobodies® are derived from the variable domain of camelide heavy chain antibodies. (b) A His-tagged Nanobody®-encoding sequence was translationally coupled to the bla gene using both the strong and weak translational coupling hairpins. (c) Nanobody® TIR libraries were constructed by PCR, transformed into BL21(DE3) pLysS and grown with different ampicillin concentrations on agar plates. (d) Colony forming unit (CFU) counts after plating the TIR libraries with the weak coupling to bla (TIR^{Lib1}) or (e) with the strong coupling to bla (TIR^{Lib2}). The growth of the original, non-randomized clones (TIR^{Orig}) were included as controls. Concentrations where optimized clones were sampled are indicated with an asterisk. (f) Nanobody® production levels with two different synthetically evolved TIR clones were analysed by Western blotting using an antibody against the His-tag. $TIR^{SynEvo1}$ was selected with the weak coupling device and $TIR^{SynEvo2}$ with the strong coupling device The increase in Nanobody® production was analysed by densitometry and plotted as the ratio of protein compared to the original clone (right panel). An antibody against Leader Peptidase (LepB) was used as a loading control. (g) The specific Nanobody® binds to GFP (illustrated in upper panel) and functionality of the nanobodies produced from the original and from the two synthetically evolved clones were assayed by mixing GFP with cell extracts from the different clones and analysed by in-gel fluorescence (lower left panel). The activity levels were estimated by densitometry comparing in-gel free GFP levels with the larger Nanobody®-bound GFP species (lower right panel).

Plasmids were recovered from five different colonies that grew at 400 µg/mL ampicillin, and five unique TIRs were identified. Protein production from these TIRs was tested in glucose-limited fed-batch fermentations, reaching cell concentrations of 200 g/kg wet weight. The resulting product concentration was increased from approx. 26 mg/g cells to 50-57 mg/g cells (Fig. 5d). Or calculated another way, from 6 g/L to 11-12 g/L, thus approximately doubling the yield. The purified product did not show any increase of product related impurities.

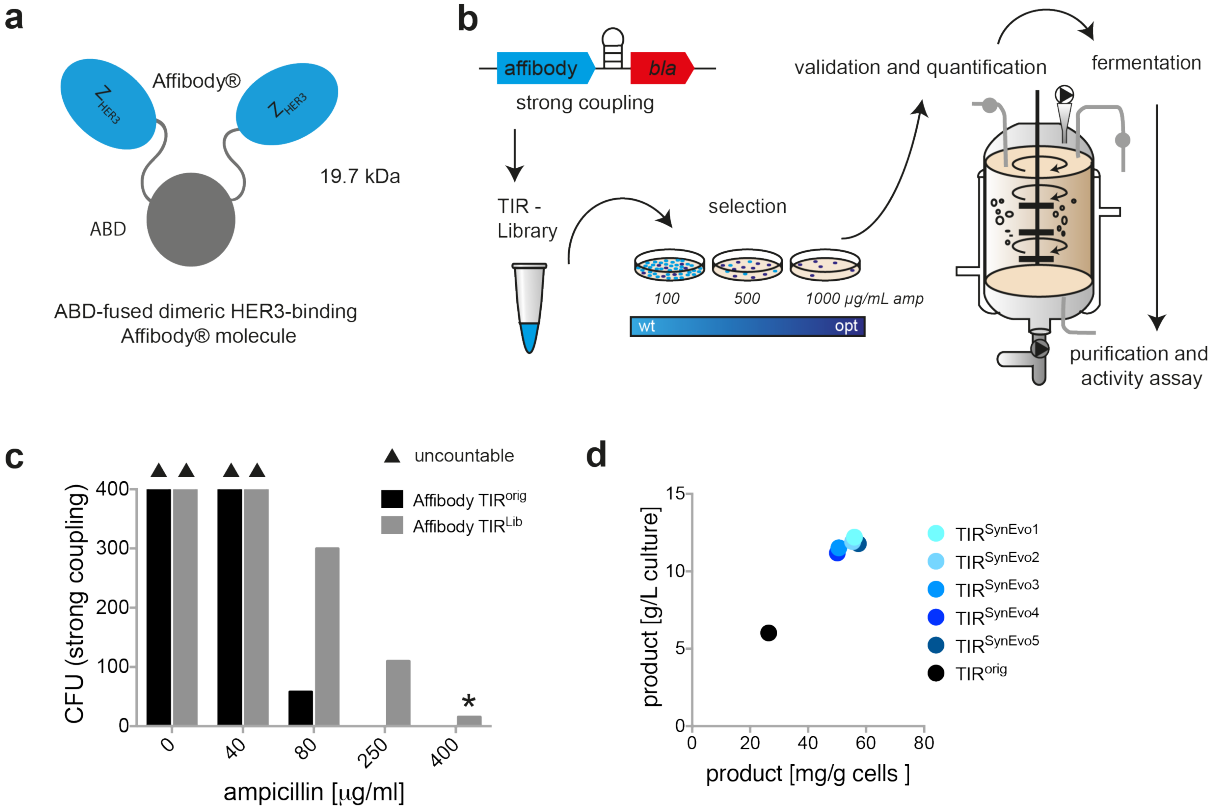


Figure 5. Optimized production of an Affibody® in fed-batch fermentation by synthetic evolution. (a) Schematic structure of the Affibody® molecule. Two Affibody® units (blue) are fused to an albumin binding domain (ABD, grey). The molecule binds bivalently to the human epidermal growth factor receptor 3 (HER3). (b) The Affibody® moelcule-encoding sequence was translationally coupled to the bla gene using a strong translational coupling hairpin. A TIR library was constructed by PCR and expressed in BL21(DE3) pLysS. Selection was performed on LB agar plates with different ampicillin concentrations. Optimized constructs were tested in fed-batch fermentation for their improved production. (c) Colony forming unit (CFU) were compared after plating the TIR library (TIR^{Lib}) and the original construct as control (TIROrig). The concentration used for sampling of optimized variants is indicated with an asterisk. (d) Affibody® production levels in fed-batch fermentation were calculated for five synthetically evolved library variants ($TIR^{SynEvo1}$ – $TIR^{SynEvo5}$) per cell mass and per culture volume in comparison to the original construct (TIR^{Orig}).

Conclusions

Previous studies from our laboratories have shown large differences in protein production by randomizing a specific region of the TIR[14]. Possibly, the approach represents a good compromise between the extend of sequence variation and library size and this enables isolation of highly

expressing clones even by simple colony picking and manual protein production screening by e.g. fluorescence or SDS-PAGE. To simplify and speed up identification of the best clones in these TIR libraries, we have now developed a selection approach that can replace manual or FACS-based screening. By translationally coupling a cds of interest to the cds of ß-lactamase and then plating the library on high ampicillin concentrations we could easily select for the best expressing gene variants. Formation of satellite colonies is a known issue with using ampicillin, but this was not observed under our experimental conditions. If this problem did occur, replacing ampicillin with the more stable carbenicillin could solve the issue[40]. Construction of the TIR library and selection of optimal clones can be performed in just a few days, and may be attractive to combine with DNA synthesis as a general synthetic biology workflow for simultaneous gene synthesis and expression optimization. The selection-based screening may also enable screening of other libraries that extend beyond the TIR, such as promoter optimization, or optimization of entire gene coding sequences (produced synthetically or by error-prone PCR).

In the present work, we evaluated the efficacy of six commonly used antibiotic resistance genes, revealing important differences in the application of those as reporters. This systematic study informs the general use of these antibiotics for any growth-based selection. Secondly, we designed novel translational-coupling devices with different coupling efficiencies. These coupling-devices can therefore expand the utility of the used antibiotic resistance genes. Finally, we designed the translational-coupling devices in such a way that no artificial extension of the protein of interest is necessary.

The six tested antibiotic resistance genes confer resistance to the most commonly used laboratory antibiotics. Five of these antibiotics encode cytosolic proteins that target the ribosome and therefore interfere with protein biosynthesis[41]. Ampicillin inhibits periplasmic proteins which are required for peptidoglycan synthesis. The protein that confers resistance, β-lactamase, is secreted into the periplasm via the Sec translocon[42]. Our analysis identified β-lactamase as the most suitable resistance marker for selection despite the requirement for secretion. This may reflect the mode of action of ampicillin, being the only tested antibiotic not acting on protein synthesis. It would be interesting to explore more antibiotics that are not acting on protein biosynthesis for this purpose since β-lactamase could saturate the secretory machinery when expressing recombinant proteins that also utilize the Sec translocon, such as membrane proteins and periplasmic proteins. Indeed, when performing whole cell fluorescence measurements for the selected high expressing NarK-GFP and AraH-GFP variants, we noticed a significant decrease in fluorescence when the

translational coupling cassette was added (on average six- and three-fold, Supplementary Fig. S2a and S2b). A similar effect was not observed when using the spectinomycin resistance gene instead of the ampicillin resistance gene in the antibiotic selection system. Therefore, it may be worth removing the selection device after identification of a high expressing gene variant, especially when working with a membrane or secreted protein.

The choice of β-lactamase as the most suitable candidate in a growth-based selection system might not be surprising. Previously β-lactamase was used as a growth-based sensor for protein stability[43,44] and very recently it was used in a survival assay that linked extracellular protein secretion to ampicillin resistance[45]. Reasons for the abrupt plateauing of the chloramphenicol resistance gene at 1 mg/mL chloramphenicol might be caused by too high ethanol concentrations in the medium (due to chloramphenicol stock solution being dissolved in 100% v/v ethanol) or is result of a limitation in the tuneable range of the system. To avoid the latter one could use a weaker variant of our translational coupling device. The general limitation of tetracycline-based selection, due to its extremely narrow dynamic range of action, may be related to the resistance gene used (*tetC*), and growth at higher concentration might be achieved by using the tetracycline transporter TetA instead. Here, it should be taken into account that it was previously observed that over-expression of a membrane proteins like the tetracycline transporter can have toxic effects on the cell 46 and can lead to impaired growth by affecting the cell's membrane potential[47].

In addition to our two model membrane proteins, we demonstrated that the translational coupling cassette facilitated the identification of optimized TIRs for the production of a Nanobody® and an Affibody®. The Nanobody® was increased from virtually no expression to an estimated 1 g/L scale yield. Previous studies reported a yield of 10-15 mg/L of culture[48], albeit under different growth conditions. With the optimization of the TIR sequence in our expression vector, the obtained yield was more than 65-fold higher. The production of the Affibody® was doubled in a fed-batch fermentation at >10 g/L industrial scale. The latter example demonstrates that even highly optimized production scenarios at industrial titres can benefit from TIR optimization and selection. Furthermore, these examples highlight the importance of the extended coupling device toolbox to enable antibiotic selection at many different production levels.

The optimization of the two pharmaceutically relevant proteins also demonstrates the strength of our developed hairpin structures in the selection system. These enable translational coupling of a gene of interest without any modification to the protein sequence, such as the addition of an

expression tag. This is particularly important for industrially- or pharmacologically-relevant proteins, where modifications to the protein sequence are not tolerable for downstream applications.

Methods

Strains, media and growth conditions

The strains used in this study are listed in Supplementary Table S1. *Escherichia coli* NEB5α (New England Biolabs, Ipswich, MA, USA) was used for cloning and propagation of plasmids. For library constructions the *E. coli* strain MC1061 was used. Expression experiments were performed in *E. coli* BL21(DE3) pLysS (Novagen, Merck KGaA, Darmstadt, Germany). *E. coli* MG1655 was used for testing combinations of RNA hairpin structures and antibiotic resistance genes. Chemically competent cells of NEB5α, MG1655 and MC1061 were obtained as described elsewhere[49]. Bacteria were cultivated in lysogeny broth (LB) with shaking (250 rpm) at 37 °C. Cultures were supplemented with antibiotics when required. Unless otherwise stated, antibiotics were used in the following concentrations: ampicillin and carbenicillin (100 μg/ml), kanamycin (50 μg/ml), chloramphenicol (34 μg/ml), spectinomycin (50 μg/ml), tetracycline (10 μg/ml) and gentamicin (10 μg/ml).

Plasmid and strain construction

All plasmid manipulations were performed using uracil excision cloning as described elsewhere50. The plasmids used in this study are listed in Supplementary Table S2. PCR was carried out with PfuX7 polymerase as previously described[51] and Phusion U Hot Start polymerase (Thermo Fisher Scientific, Waltham, MA, USA). Oligonucleotides were received from Integrated DNA technologies (IDT, Coralville, IA, USA) (Supplementary Table S3). For comparing different translationally-coupled antibiotic resistance genes and different RNA hairpin structures, the pACYC-DasherGFP-hp-AbR series was constructed using the resistance genes of different vector backbones (Supplementary Table S2) and a plasmid encoding the DasherGFP protein under control of a rhamnose inducible promoter (ATUM Bio, Newark, CA, USA). The various hairpin sequences were introduced by PCR with oligonucleotides. For validation of the selection system the hp-*bla* and hp-*aadA* module were cloned into a pET28a(+) backbone already encoding known expression variants of AraH-GFP and NarK-GFP. For optimization of the GFP-Nanobody® the gene encoding the Nanobody®, as well as translational coupling element and bla gene, were cloned into the pET28a(+) backbone (Novagen, Merck KGaA, Darmstadt, Germany). All plasmids

were isolated using the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany) and PCR products were purified using the NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel, Düren, Germany).

Library design and construction

TIR libraries were constructed using a degenerated forward primer specific for each gene of interest. In each degenerated forward primer the six nucleotides upstream of the start codon were changed to all possible combinations and the six nucleotides downstream of the start codon were changed to all possible synonymous codon combinations. For plasmid library constructions Q5 polymerase was used (New England Biolabs, Ipswich, MA, USA). Library construction was performed as described elsewhere[14].

Expression and selection

For expression of individual clones, overnight cultures were prepared by inoculating a single colony in 800 µL LB media supplemented with respective antibiotics and incubated in a 2.2 ml 96-deep well plate (EnzyScreen, Heemstede, The Netherlands) at 37 °C with shaking. Cultures were then back-diluted (1:50) into 5 mL of LB media containing appropriate antibiotics in a 24-deep well plate and incubated at 37 °C. At an OD_{600} of approximately 0.3 expression was induced with 1.0 mM IPTG. For membrane protein expression the temperature was lowered to 25 °C.

For expression and selection from clone libraries ca. 1 µg of plasmid library was transformed into 100 µl BL21(DE3)pLysS using a standard protocol. The transformation was transferred to 5 ml LB media supplemented with respective antibiotics and grown overnight at 37 °C. Cultures were then back-diluted (1:50) into 5 mL LB media containing the appropriate antibiotics in a 24-deep well plate (EnzyScreen, Heemstede, The Netherlands) and incubated at 37 °C. At an OD_{600} of approximately 0.3 expression was induced with 1.0 mM IPTG. For membrane protein expression the temperature was lowered to 25 °C. 2 h after induction, OD_{600} was measured and 0.2 OD units of cells were then plated on LB agar plates containing 1 mM IPTG and different concentrations of ampicillin and incubated overnight at 37 °C. Selected expression variants were sequenced. To exclude other reasons than an optimized translation initiation region for higher expression levels, the selected TIR sequences were re-cloned into a clean expression vector for further characterization.

MIC determination

To test translationally coupled antibiotic resistance genes and different hairpins, the pACYC-DasherGFP-hp-AbR series were transformed into *E. coli* MG1655 using a standard protocol. Overnight cultures were prepared by inoculating a single colony in 3 mL 2xYT media containing the appropriate antibiotic and incubated at 37 °C. Cultures were then back-diluted (1:100) into 2 mL 2xYT media containing the appropriate antibiotic and different concentrations of rhamnose as inducer (0 mM, 0.625 mM, 1.25 mM, 2.5 mM, 5 mM, 10 mM, 20 mM, 40 mM rhamnose) in a 24-deep well plate (EnzyScreen, Heemstede, The Netherlands). Cultures were incubated at 37 °C with shaking. After 4 h, 10 μL of each culture was transferred to a 96 well plate (Greiner, Kremsmünster, Austria) containing 100 μL 2xYT media (ca. 5x105 cfu/mL), a serial dilution of the antibiotic and respective concentrations of rhamnose as inducer. Plates were incubated for 18 h. For end-point MIC determination fluorescence (Ex: 485 nm, Em: 512 nm for GFP; Ex:, 512 nm Em: 520 nm for DasherGFP) and OD_{600} were measured in an MX plate reader (Biotek, Winooski, VT, USA). A culture containing none of the inducible antibiotic served as positive control, media containing no culture served as negative control. For each antibiotic the MIC value was defined as the lowest antibiotic concentration at which the final OD_{600} represented less than 10% of the entire population after background correction. Each MIC experiment was conducted with biological triplicates.

To assess protein production, constructs were transformed into *E. coli* BL21(DE3) pLysS (Novagen, Merck KGaA, Darmstadt, Germany) using a standard protocol. Cultures were prepared and expression was performed as described above. After induction, cultures were incubated for 2 h. For MIC determination a dilution series of the translationally coupled antibiotic was prepared in a 96 well plate (Greiner, Kremsmünster, Austria). Subsequently, 10 μL of each culture was transferred to the 96 well plate.

Flow cytometry

For flow cytometry measurements, cultures were prepared and expression was performed as described above. Expression cultures were grown overnight after induction and then back-diluted (1:100) in 1x PBS for analysis. Flow cytometry measurements were performed on a FACS Aria (Becton–Dickinson, San Jose, CA, USA) with excitation at 488 nm from a blue solid-state laser. FlowJo (Treestar, Ashland, OR, USA) was used for data analysis.

Western Blot analysis

For Western Blot analysis, cultures were prepared and expression was performed as described above. Cultures were grown for 5 h after induction. Cells were resuspended to a final concentration of 0.025 ODU/µl in CelLytic B (Sigma Aldrich, St. Louis, MO, USA) supplemented with lysozyme, egg white (Amresco, Solon, OH, USA), Benzonase nuclease (≥250 units/µL, Sigma Aldrich, St. Louis, MO, USA) and Roche cOmplete™ Protease Inhibitor Cocktail (Sigma Aldrich, St. Louis, MO, USA) and incubated for 1 h before the sample was mixed with the same volume of 5x reducing sample buffer and heated to 95°C for 5 min for protein denaturation. 0.05 ODU of the samples were loaded onto a 4-20 % Mini-PROTEAN-TGX gel (BioRad, Hercules, CA, USA) and run for 35 min at 175 V. Proteins were transferred to a nitrocellulose membrane using the iBlot® dry blotting system (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) at 25 V for 7 min. The proteins were detected using antigen-specific antibodies. The following antibodies were used: anti-His (1:1000; Merck Millipore, Merck KGaA, Darmstadt, Germany), anti-Lep (1:1000; a generous gift from IngMarie Nilsson and Gunnar von Heijne, Stockholm University). Primary antibodies were diluted in 5% w/v skim milk in TBS-T (20 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1 % v/v Tween-20), secondary antibodies were diluted in TBS-T. The secondary antibody was visualized using Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare, Chicago, IL, USA). The chemoluminescence signal was detected using G:Box bioimager (Syngene, Cambridge, UK). The resulting images were analyzed by densitometry using the Fiji software[52].

Analyses of expression levels of Nanobody® molecules

Cell pellets of 10 mL cultures were resuspended in 600 µL CelLytic B (Sigma Aldrich, St. Louis, MO, USA) supplemented with lysozyme, egg white (Amresco, Solon, OH, USA), Benzonase nuclease (≥250 units/µL, Sigma Aldrich, St. Louis, MO, USA) and Roche cOmplete™ Protease Inhibitor Cocktail (Sigma Aldrich, St. Louis, MO, USA) and incubated for 1 h on ice. Lysates were centrifuged for 20 min at 12,000 x g and the soluble fraction was loaded on a Ni-NTA Spin Column (Qiagen, Hilden, Germany). The spin column was washed twice with wash buffer (600 µL of 50 mM NaH2PO4, pH 7.4, 300 mM NaCl and 20 mM imidazole) and then eluted twice with 300 µl of elution buffer (50 mM Tris-Cl pH 7.4, 300 mM NaCl, 250 mM imidazole). The two eluates were pooled. Protein concentration in the 600 uL eluate were estimated by densiometric analysis in accordance to a standard curve of known protein concentration using the Fiji software[52].

In-Gel fluorescence measurements

For In-Gel fluorescence measurements, cultures were prepared and expression was performed as described above. Cultures were grown for 5 h after induction. Cells were resuspended to a final concentration of 0.05 ODU/µl in CelLytic B (Sigma Aldrich, St. Louis, MO, USA) supplemented with lysozyme, egg white (Amresco, Solon, OH, USA), Benzonase nuclease (≥250 units/µL, Sigma Aldrich, St. Louis, MO, USA) and Roche cOmplete™ Protease Inhibitor Cocktail (Sigma Aldrich, St. Louis, MO, USA) buffer and incubated for 20 min on ice. Subsequently, cells were mixed with 0.2 mg/mL GFP (final concentration 0.1 mg/ml). Cells and GFP were incubated for 20 min at 30 °C and 250 rpm. 10 µl of the GFP-cell-mixture were mixed with 5 µl 2× Laemmli sample buffer and the whole sample was loaded onto a 4-20 % Mini-PROTEAN-TGX gel (BioRad, Hercules, CA, USA) and run for 35 min at 175 V. The fluorescence signal was detected using G:Box bioimager (Syngene, Cambridge, UK). The resulting images were analysed by densitometry using the Fiji software[52].

Expression of Affibody® molecules

Pre-inoculi were made from fresh colonies on selective agar using TSB+YE-medium supplemented with 50 mg/mL kanamycin. After 5 h incubation at 37 °C part of the pre-inoculi were used to inoculate a YNB-based defined shake-flask medium. This culture was incubated at 30 °C for 22 h. A defined synthetic fermentor medium was inoculated to an OD of 0.112. The fermentations were run at 37 °C using a glucose feeding strategy to control the growth rate, using a Greta Multifermentor System (Belach Bioteknik AB, Skogås, Sweden). pH was controlled at 7, through the automatic titration with 25 % NH4OH. After 18 h, the fermentations were induced with 0.5 mM IPTG. The fermentations were harvested after 30 h total cultivation time. Cell pellet samples were collected through centrifugation.

Analyses of expression levels and purity of Affibody® molecules

Cell pellets were treated with CelLytic (Sigma, St. Louis, MO, USA) and purified on product specific affinity chromatography matrices. The absorbance at 280 nm of the eluate was measured, and the product concentration could be calculated. The purity was analysed using an LC/MS method.

Abbreviations

aadA	gene encoding streptomycin 3''-adenylyltransferase
ABD	albumin binding domain
acc3	gene encoding aminoglycoside N(3)-acetyltransferase
aphA1	gene encoding aminoglycoside 3'-phosphotransferase
bla	gene encoding β-lactamase
cat	gene encoding chloramphenicol acetyltransferase
cds	coding sequence
FACS	fluorescence-activated cell sorting
GFP	green fluorescent protein
HER3	human epidermal growth factor receptor 3
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	lysogeny broth
MIC	minimum inhibitory concentration
PBS	phosphate buffered saline
PCR	polymerase chain reaction
rRNA	ribosomal RNA
SD	Shine Dalgarno
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
tetC	gene encoding TetC, tetracycline resistance protein, class C
TIR	translation initiation region
TSB	tryptic soy broth
YE	yeast extract
YNB	yeast nitrogen base

Author Information

Corresponding Author

*Tel: +45 217-99184. Fax: +45-353-33300. E-mail: morno@biosustain.dtu.dk.

Author contributions

MR and MHHN designed the study. MR, VM, FD and KM performed the experiments. DOD, FD and MHHN supervised the research. The manuscript was prepared by MR and MHHN with contributions from all authors.

Funding

This work was supported by The Novo Nordisk Foundation (MHHN) and the Swedish Research Council (DOD). It is noted that the subject matter of this publication forms the basis of a patent application number filed by co-inventors MR, VM, KM, DOD and MHHN.

Acknowledgments

The authors would like to acknowledge Camilla Johansen, Josephine Roien and Sofia Carlsson Tysk for their help with the GFP-Nanobody® library.

References

1. Rosano, G. L., and Ceccarelli, E. A. (2014) Recombinant protein expression in Escherichia coli : advances and challenges. Front. Microbiol. 5, 1–17.
2. Terpe, K. (2006) Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. Appl. Microbiol. Biotechnol. 72, 211–222.
3. Miroux, B., and Walker, J. E. (1996) Over-production of Proteins inEscherichia coli: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. J. Mol. Biol. 260, 289–298.
4. Hammarström, Martin, Hellgren, Niklas, van den Berg, Susanne, Berglund, Helena and Härd, T. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Sci. 11, 313–321.
5. Kapust, R. B., and Waugh, D. S. (1999) Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci. 8, 1668–1674.
6. Baker, R. T. (1996) Protein expression using ubiquitin fusion and cleavage. Curr. Opin. Biotechnol. 7, 541–546.
7. Smith, D. B., and Johnson, K. S. (1988) Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene 67, 31–40.
8. Terpe, K. (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl. Microbiol. Biotechnol. 60, 523–533.
9. Laursen, B. S., Sørensen, H. P., Mortensen, K. K., and Sperling-Petersen, H. U. (2005) Initiation of Protein Synthesis in Bacteria. Microbiol. Mol. Biol. Rev. 69, 101–123.
10. Kozak, M. (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene 361, 13–37.

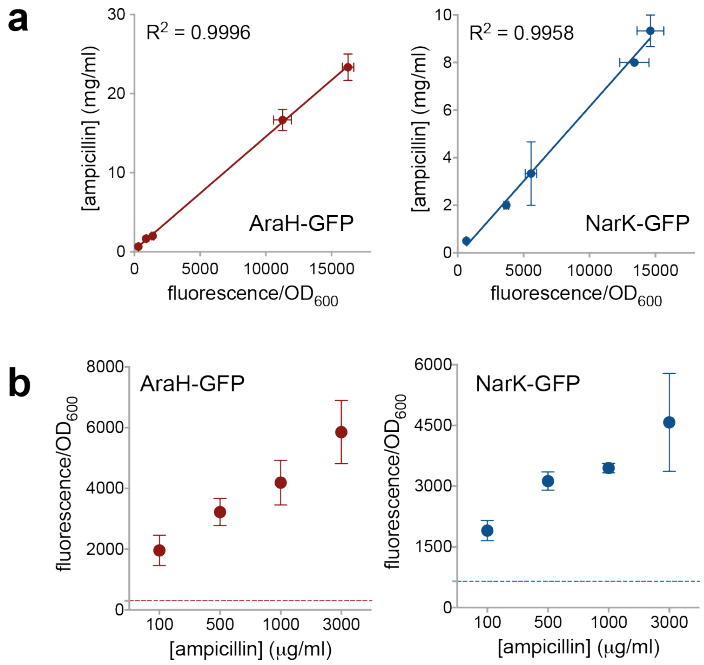
11. Gold, L. (1988) Posttransriptional regulatory mechanisms in Escherichia coli. Annu. Rev. Biochem. 57, 199–233.
12. de Smit, M. H., and van Duin, J. (2003) Translational Standby Sites : How Ribosomes May Deal with the Rapid Folding Kinetics of mRNA. J. Mol. Biol. 2836, 737–743.
13. Reeve, B., Hargest, T., Gilbert, C., and Ellis, T. (2014) Predicting translation initiation rates for designing synthetic biology. Front. Bioeng. Biotechnol. 2.
14. Mirzadeh, K., Martínez, V., Toddo, S., Guntur, S., Herrgård, M. J., Elofsson, A., Nørholm, M. H. H., and Daley, D. O. (2015) Enhanced Protein Production in Escherichia coli by Optimization of Cloning Scars at the Vector–Coding Sequence Junction. ACS Synth. Biol. 4, 959–965.
15. Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., and Endy, D. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. Nat. Methods 10, 354–360.
16. Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., and Blüthgen, N. (2013) Efficient translation initiation dictates codon usage at gene start. Mol. Syst. Biol. 9, 1–10.
17. Kudla, G., Murray, A. W., Tollervey, D., and Plotkin, J. B. (2009) Coding-sequence determinants of gene expression in Escherichia coli. Science. 324, 255–258.
18. Goodman, D. B., Church, G. M., and Kosuri, S. (2013) Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. Science. 342, 475–479.
19. Plotkin, J. B., and Kudla, G. (2011) Synonymous but not the same: the causes and consequences of codon bias. Nat. Rev. Genet. 12, 32–42.
20. Mortimer, S. A., Kidwell, M. A., and Doudna, J. A. (2014) Insights into RNA structure and function from genome-wide studies. Nat. Rev. Genet. 15, 469–479.
21. Bonde, M. T., Pedersen, M., Klausen, M. S., Jensen, S. I., Wulff, T., Harrison, S., Nielsen, A. T., Herrgård, M. J., and Sommer, M. O. A. (2016) Predictable tuning of protein expression in bacteria. Nat. Methods 13, 233–236.
22. Na, D., Lee, S., and Lee, D. (2010) Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with desired expression levels in prokaryotes. BMC Syst. Biol. 4.
23. Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. Nat. Biotechnol. 27, 946–950.
24. Woo, S., Yang, J., Kim, I., Yang, J., Eun, B., Kim, S., and Yeol, G. (2013) Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. Metab. Eng. 15, 67–74.
25. Borujeni, A. E., Channarasappa, A. S., and Salis, H. M. (2013) Translation rate is controlled by coupled trade-offs between site accessibility , selective RNA unfolding and sliding at upstream standby sites. Nucleic Acids Res. 42, 2646–2659.
26. Tian, T., and Salis, H. M. (2015) A predictive biophysical model of translational coupling to coordinate and control protein expression in bacterial operons. Nucleic Acids Res. 43, 7137–7151.

27. Skretas, G., and Georgiou, G. (2010) Simple Genetic Selection Protocol for Isolation of Overexpressed Genes That Enhance Accumulation of Membrane-Integrated Human G Protein-Coupled Receptors in Escherichia coli. Appl. Environ. Microbiol. 76, 5852–5859.
28. Massey-Gendel, E., Zhao, A., Boulting, G., Kim, H.-Y., Balamotis, M. A., Seligman, L. M., Nakamoto, R. K., and Bowie, J. U. (2009) Genetic selection system for improving recombinant membrane protein expression in E. coli. Protein Sci. 18, 372–383.
29. Gul, N., Linares, D. M., Ho, F. Y., and Poolman, B. (2014) Evolved Escherichia coli strains for amplified, functional expression of membrane proteins. J. Mol. Biol. 426, 136–149.
30. Mendez-Perez, D., Gunasekaran, S., Orler, V. J., and Pfleger, B. F. (2012) A translation-coupling DNA cassette for monitoring protein translation in Escherichia coli. Metab. Eng. 14, 298–305.
31. Rex, G., Surin, B., Besse, G., Schneppe, B., and John E.G., M. (1994) The Mechanism of Translational Coupling in Escherichia coli. J. Biol. Chem. 269, 18118–18127.
32. Poole, E. S., Brown, C. M., and Tate, W. P. (1995) The identity of the base following the stop codon determines the efficiency of in vivo translational termination in Escherichia coli. EMBO J. 14, 151–8.
33. Chen, H., Bjerknes, M., Kumar, R., and Jay, E. (1994) Determination of the optimal aligned spacing between the shine - dalgarno sequence and the translation initiation codon of escherichia coli m RNAs. Nucleic Acids Res. 22, 4953–4957.
34. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.
35. Nørholm, M. H. H., Toddo, S., Virkki, M. T. I., Light, S., von Heijne, G., and Daley, D. O. (2013) Improved production of membrane proteins in Escherichia coli by selective codon substitutions. FEBS Lett. 587, 2352–2358.
36. Kirchhofer, A., Helma, J., Schmidthals, K., Frauer, C., Cui, S., Karcher, A., Pellis, M., Muyldermans, S., Casas-Delucchi, C. S., Cardoso, M. C., Leonhardt, H., Hopfner, K.-P., and Rothbauer, U. (2010) Modulation of protein properties in living cells using nanobodies. Nat. Struct. Mol. Biol. 17, 133–8.
37. Kubala, M. H., Kovtun, O., Alexandrov, K., and Collins, B. M. (2010) Structural and thermodynamic analysis of the GFP:GFP-nanobody complex. Protein Sci. 19, 2389–2401.
38. Frejd, F. Y., and Kim, K. T. (2017) Affibody molecules as engineered protein drugs. Exp. Mol. Med. 49, e306.
39. Bass, T. Z., Rosestedt, M., Mitran, B., Frejd, F. Y., Löfblom, J., Tolmachev, V., Ståhl, S., and Orlova, A. (2017) In vivo evaluation of a novel format of a bivalent HER3-targeting and albumin-binding therapeutic affibody construct. Sci. Rep. 7, 43118.
40. Sambrook, J. (2001) Molecular Cloning: A Laboratory Manual. 3rd ed. Cold Spring Harbor, N.Y. :Cold Spring Harbor Laboratory Press.
41. Wilson, D. N. (2014) Ribosome-targeting antibiotics and mechanisms of bacterial resistance. Nat. Rev. Microbiol. 12, 35–48.
42. Pradel, N., Delmas, J., Wu, L. F., Santini, C. L., and Bonnet, R. (2009) Sec- and Tat-dependent translocation of β-lactamases across the Escherichia coli inner membrane. Antimicrob. Agents Chemother. 53, 242–248.

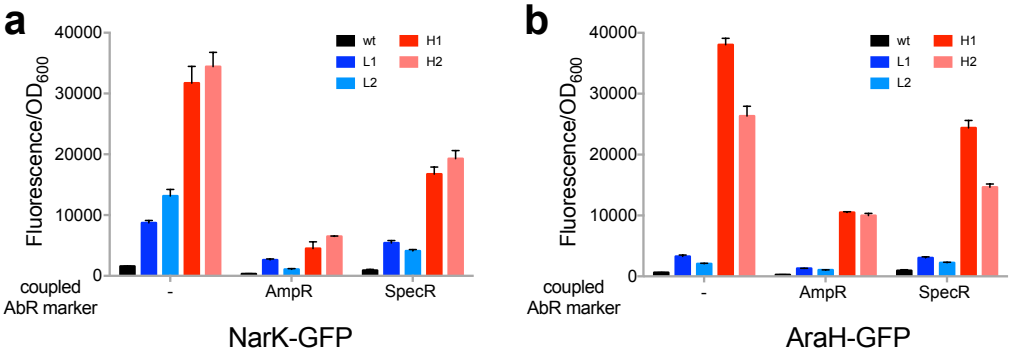
43. Foit, L., Morgan, G. J., Kern, M. J., Steimer, L. R., von Hacht, A. A., Titchmarsh, J., Warriner, S. L., Radford, S. E., and Bardwell, J. C. A. (2009) Optimizing Protein Stability In Vivo. Mol. Cell 36, 861–871.
44. Quan, S., Koldewey, P., Tapley, T., Kirsch, N., Ruane, K. M., Pfizenmaier, J., Shi, R., Hofmann, S., Foit, L., Ren, G., Jakob, U., Xu, Z., Cygler, M., and Bardwell, J. C. a. (2011) Genetic selection designed to stabilize proteins uncovers a chaperone called Spy. Nat. Struct. Mol. Biol. 18, 262–269.
45. Natarajan, A., Haitjema, C. H., Lee, R., Boock, J. T., and DeLisa, M. P. (2017) An Engineered Survival-Selection Assay for Extracellular Protein Expression Uncovers Hypersecretory Phenotypes in Escherichia coli. ACS Synth. Biol. 6, 875–883.
46. Guay, G. G., and Rothstein, D. M. (1993) Expression of the tetK gene from Staphylococcus aureus in Escherichia coli : Comparison of Substrate Specificities of TetA (B), TetA (C), and TetK Efflux Proteins. Antimicrob. Agents Chemother. 37, 191–198.
47. Valenzuela, M. S., Siddiqui, K. a, and Sarkar, B. L. (1996) High expression of plasmid-encoded tetracycline resistance gene in E. coli causes a decrease in membrane-bound ATPase activity. Plasmid 36, 19–25.
48. Rothbauer, U., Zolghadr, K., Muyldermans, S., Schepers, A., Cardoso, M. C., and Leonhardt, H. (2008) A Versatile Nanotrap for Biochemical and Functional Studies with Fluorescent Fusion Proteins. Mol. Cell. Proteomics 7, 282–289.
49. Inoue, H., Nojima, H., and Okayama, H. (1990) High efficiency transformation of Escherichia coli with plasmids. Gene 96, 23–28.
50. Cavaleiro, A. M., Kim, S. H., Seppälä, S., Nielsen, M. T., and Nørholm, M. H. H. (2015) Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning. ACS Synth. Biol. 4, 1042–1046.
51. Nørholm, M. H. H. (2010) A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. BMC Biotechnol. 10:21.
52. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., and Cardona, A. (2012) Fiji : an open-source platform for biological-image analysis. Nat. Methods 9, 676–682.

Supplementary information for:

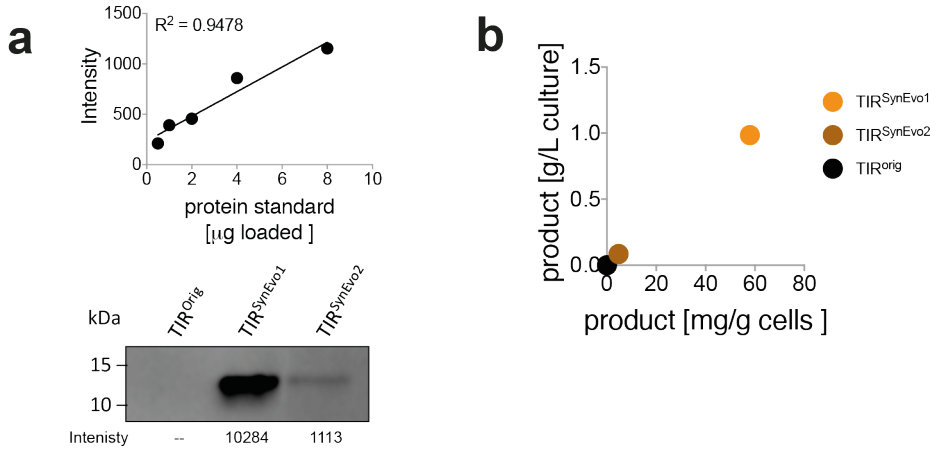
TARSyn: Tuneable antibiotic resistance devices enabling bacterial synthetic evolution and protein production



Supplementary Figure S1. The β-lactamase gene can be used as a sensor for high protein production levels. (a) A translational coupling device and the ampicillin resistance conferring gene (bla, red) were cloned into a pET28a(+) vector encoding NarK-GFP and AraH-GFP. (b) Known expression variants of the translational coupled NarK-GFP and AraH-GFP constructs were tested in minimum inhibitory concentration (MIC) assays and compared to non-coupled NarK-GFP and AraH-GFP expression levels, estimated by fluorescence. (C) Translationally coupled NarK-GFP and AraH-GFP TIR libraries grown with different ampicillin concentrations on agar plates. Ten individual colonies were picked from each plate and expression levels were estimated.



Supplementary Figure S2. Assessment of effects of the co-production of an antibiotic resistance marker. *narK-gfp* (a) and *araH-gfp* (b) expression levels were analysed in regards to co-expression of the translationally coupled ampicillin resistance conferring gene *bla* (AmpR marker) and the spectinomycin resistance conferring gene *aadA1* (SpecR marker). The wild type (wt; black), two low expressing (L1, L2; red) and two high expressing (H1, H2; blue) variants from a TIR library in a pET28a(+) backbone (KanR) were analysed.



Supplementary Figure S3. Assessment of Nanobody® expression levels. (a) A standard curve (upper panel) was used to estimate concentration of purified Nanobody® protein by densitometry (lower panel). (b) Nanobody® production levels were calculated for the two synthetically evolved library variants ($TIR^{SynEvo1}$ and $TIR^{SynEvo2}$) in comparison to the original construct (TIR^{Orig}) and plotted as product per cell mass against product per culture volume.

Table S1. Strains used in this study

Strain	Genotype	Source/Reference
E. coli NEB5α	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	*a*
E. coli MC1061	araD139, Δ(ara, leu)7697, ΔlacX74, galU-, galK-, hsr-, hsm+, strA	In house
E. coli BL21(DE3) pLysS	F^- *ompT* $hsdS_B(r_B^- m_B^-)$ *gal dcm* (DE3) pLysS (Cm^R)	*b*
E. coli MG1655	K-12 $F^- \lambda^-$ *ilvG*$^-$ *rfb-50 rph-1*	*c*

[a]NEB, Ipswich, MA, USA; [b] Novagen, Merck KGaA, Darmstadt, Germany; [c]ATCC, Manassas, VA, USA

Table S2. Plasmids used in this study

Plasmid	Property	Source/Reference
DasherGFP	vector encoding *dasher*, Km^R	*d*
pACYCDuet-1	Cloning and expression vector, Cm^R	*b*
pCDF-Duet	Cloning and expression vector, Sp^R	*b*
pET PfuX7	Cloning and expression vector, Amp^R	4
pSEVA551	Cloning and expression vector, Tet^R	5
pSEVA614	Cloning and expression vector, Gm^R	5
pET28a(+)	Cloning and expression vector, Km^R	*b*
pACYC-pRha-*dasher*-hp-*bla*	Cloning and expression vector with rhamnose promoter, CmR, p15A origin	This study
pACYC-pRha-*dasher*-hp-*aphA1*	Cloning and expression vector, Cm^R	This study
pACYC-pRha-*dasher*-hp-*aadA*	Cloning and expression vector, Cm^R	This study
pACYC-pRha-*dasher*-hp-*cat3*	Cloning and expression vector, Amp^R	This study
pACYC-pRha-*dasher*-hp-*acc3*	Cloning and expression vector, Cm^R	This study

pACYC-pRha-*dasher*-hp-*tetA*	Cloning and expression vector, Cm^R	This study
pACYC-pRha-*dasher*-hpAUU-*bla*	Cloning and expression vector, Cm^R	This study
pACYC-pRha-*dasher*-hpAUU-SD2-*bla*	Cloning and expression vector, Cm^R	This study
pET28a-*narK*^{WT}	*narK-gfp-tev-his8* with T7 promoter, Km^R, pBR322/ROP origin	6
pET28a-*narK*$^{P3-3}$	*narK-gfp-tev-his8* with T7 promoter, Km^R, pBR322/ROP origin	6
pET28a-*narK*$^{P4-4}$	*narK-gfp-tev-his8* with T7 promoter, Km^R, pBR322/ROP origin	6
pET28a-*narK*$^{P7-1}$	*narK-gfp-tev-his8* with T7 promoter, Km^R, pBR322/ROP origin	6
pET28a-*narK*$^{P7-6}$	*narK-gfp-tev-his8* with T7 promoter, Km^R, pBR322/ROP origin	6
pET28a-*araH*^{WT}	*araH-gfp-tev-his8* with T7 promoter, Km^R, pBR322/ROP origin	6
pET28a-*araH*$^{P2-2}$	*araH-gfp-tev-his8* with T7 promoter, Km^R, pBR322/ROP origin	6
pET28a-*araH*$^{P2-3}$	*araH-gfp-tev-his8* with T7 promoter, Km^R, pBR322/ROP origin	6
pET28a-*araH*$^{P7-4}$	*araH-gfp-tev-his8* with T7 promoter, Km^R, pBR322/ROP origin	6
pET28a-*araH*^{opt}	*araH-gfp-tev-his8* with T7 promoter, Km^R, pBR322/ROP origin	7
pET28a-*narK*^{WT}-*bla*	*narK-gfp-tev-his8*-hp-*bla* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*narK*$^{P3-3}$ -*bla*	*narK-gfp-tev-his8*-hp-*bla* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*narK*$^{P4-4}$-*bla*	*narK-gfp-tev-his8*-hp-*bla* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*narK*$^{P7-1}$-*bla*	*narK-gfp-tev-his8*-hp-*bla* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*narK*$^{P7-6}$-*bla*	*narK-gfp-tev-his8*-hp-*bla* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*araH*^{WT}-*bla*	*araH-gfp-tev-his8*-hp-*bla* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*araH*$^{P2-2}$-*bla*	*araH-gfp-tev-his8*-hp-*bla* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*araH*$^{P2-3}$-*bla*	*araH-gfp-tev-his8*-hp-*bla* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*araH*$^{P7-4}$-*bla*	*araH-gfp-tev-his8*-hp-*bla* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*araH*^{opt}-*bla*	*araH-gfp-tev-his8*-hp-*bla* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*narK*^{WT}-*aadA*	*narK-gfp-tev-his8*-hp-*aadA* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*narK*$^{P3-3}$-*aadA*	*narK-gfp-tev-his8*-hp-*aadA* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*narK*$^{P4-4}$-*aadA*	*narK-gfp-tev-his8*-hp-*aadA* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*narK*$^{P7-1}$-*aadA*	*narK-gfp-tev-his8*-hp-*aadA* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*narK*$^{P7-6}$-*aadA*	*narK-gfp-tev-his8*-hp-*aadA* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*araH*^{WT}-*aadA*	*araH-gfp-tev-his8*-hp-*aadA* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*araH*$^{P2-2}$-*aadA*	*araH-gfp-tev-his8*-hp-*aadA* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*araH*$^{P2-3}$-*aadA*	*araH-gfp-tev-his8*-hp-*aadA* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*araH*$^{P7-4}$-*aadA*	*araH-gfp-tev-his8*-hp-*aadA* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*araH*^{opt}-*aadA*	*araH-gfp-tev-his8*-hp-*aadA* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*Nanobody*®-hp-*bla*	*Nanobody*®-*his6*-hp-*bla* with T7 promoter, Km^R, pBR322/ROP origin	This study

pET28a-*Nanobody*®-hpAUU-*bla*	*Nanobody*®-*his6*-hpAUU-*bla* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*Nanobody*®-$TIR^{SynEvo1}$	*SynEvo1-Nanobody*®-*his6* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*Nanobody*®-$TIR^{SynEvo2}$	*SynEvo2-Nanobody*®-*his6* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*Nanobody*®-TIR^{BCD}	*BCD-Nanobody*®-*his6* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*Nanobody*®-TIR^{RBSopt}	*RBSopt-Nanobody*®-*his6* with T7 promoter, Km^R, pBR322/ROP origin	This study
pET28a-*Nanobody*®-TIR^{UTRopt}	*UTRopt-Nanobody*®-*his6* with T7 promoter, Km^R, pBR322/ROP origin	This study

[b]Novagen, Merck KGaA, Darmstadt, Germany; [d]ATUM, Newark, CA, USA

Table 3. Oligonucleotides used in this study

Name	Sequence (5' --> 3')
Cloning oligonucleotides	
pET28a-backbone_fwd	ATC CGG CUG CTA ACA AAG CCC GAA AG
pET28a-backbone_rev	ATC CTC GAG UCT CCT TCT TAA AG
Amp-pET28a_rev	AGC CGG AUC TCA TTA CCA ATG CTT AAT C
Spec-pET28a_rev	AGC CGG AUC TCA TTA TTT GCC GAC TAC CTT GGT GAT CTC G
pACYC_backbone_rev	AGG TAT CCU CAG CCG CGC GCG CGT C
pACYC_backbone_fwd	AGG TTA CCU CAG CGG CCG GCC CCT G
pRhaDasher_fwd	AGG ATA CCU CAG CCA CCA CAA TTC AGC AAA TTG TG
pRhaDasher_hp_rev	ACC TCC TAU GTC ACT GAT ACG TGT CCA GAT CAA CCG C
pRhaDasher_hpAUU_rev	ACC TCC TAU TTC ACT GAT ACG TGT CCA GAT CAA CCG C
pRhaDasher_hpAUU_SD2_rev	ACC AGG AAU TTC ACT GAT ACG TGT CCA GAT CAA CCG C
Amp_hp_fwd	ATA GGA GGU CCT CCT ATG TCA ATT CAA CAT TTC CGT GTC
Amp_hpAUU_fwd	ATA GGA GGU CCT CCT ATT TCA ATT CAA CAT TTC CGT GTC
Amp_hpAUU_SD2_fwd	ATT CCT GGU CCA GGA ATT TCA ATT CAA CAT TTC CGT GTC
Spec_hp_fwd	ATA GGA GGU CCT CCT ATG TCA AGG GAA GCG GTG ATC GC
Kan_hp_fwd	ATA GGA GGU CCT CCT ATG TCA CAT ATT CAA CGG GAA AC
Cm_hp_fwd	ATA GGA GGU CCT CCT ATG TCA GAG AAA AAA ATC ACT GG
Tet_hp_fwd	ATA GGA GGU CCT CCT ATG TCA AAA TCT AAC AAT GCG CTC
Gm_hp_fwd	ATA GGA GGU CCT CCT ATG TCA TTA CGC AGC AGC AAC
Tet_ rev	AGG TAA CCU CAG CTC AGG TCG AGG TGG CCC GGC TC
Cm_ rev	AGG TAA CCU CAG CTT ACG CCC CGC CCT GCC AC
Kan_ rev	AGG TAA CCU CAG CTT AGA AAA ACT CAT CGA GCA TCA AAT G
Spec_ rev	AGG TAA CCU CAG CTT ATT TGC CGA CTA CCT TGG TGA TCT C
Gm_ rev	AGG TAA CCU CAG CTT AGG TGG CGG TAC TTG GGT C
Nanobody®_fwd	ACT CGA GGA UGG CTC AGG TCC AAC
Nanobody®_hpAUU_rev	ACC TCC TAU TTC AGT GGT GGT GGT GG
Nanobody®_hp_rev	ACC TCC TAU GTC AGT GGT GGT GGT GG
Nanobody®_var1_fwd	ACT TTA AGA AGG AGA CTG GTA AAT GGC CCA AGT CCA ACT GGT CGA ATC AG
Nanobody®_var2_fwd	ACT TTA AGA AGG AGA CGA ATA TAT GGC TCA AGT CCA ACT GGT CGA ATC AG
Nanobody®_RBSopt_fwd	ATT CCC CTC UAG AAT AAT AGA GAC GTA ATA AGT AAT AAG GAG GTA AAC ATG GCT CAG GTC CAA CTG GTC GAA TCA G
Nanobody®_UTRopt_fwd	ATT CCC CTC UAG AAT AAT GTC ACA ATG CAA AGG AGA ATA TAC AAT GGC TCA GGT CCA ACT GGT CGA ATC AG
Nanobody®_RBS_rev	AGA GGG GAA UTG TTA TCC GCT CAC AAT TCC CCT ATA GTG AGT CGT ATT A
Nanobody®_BCD_fwd	AAA CAT CTT AAU CAT GCT AAG GAG GTT TTC TAA TGG CTC AGG TCC AAC TGG TCG AAT C
Nanobody®_BCD_rev	ATT AAG ATG TTU CAG TAC GAA AAT TGC TTT CAT CCT CGA GTC TCC TTC TTA AAG TTA AAC
Bla_loopout_fwd	ATA GGA GGU TGA GAT CCG GCT GCT AAC AAA G

Sequencing oligonucleotides	
hp-Amp_seq_fwd	GAC AAC GAT CGG AGG ACC GAA G
hp-Amp_seq_rev	GAT CAA GGC GAG TTA CAT GAT C
T7P_seq_fwd	TAA TAC GAC TCA CTA TAG GGG AAT TG
Dasher-hp_seq_fwd	CAT CCG CGT TGA GTT CAA CC
T7T_seq_rev	GCT CAG CGG TGG CAG CAG CCA ACT CAG CTT
Duet_colPCR_fwd	GCG ACT CCT GCA TTA GGA AA
Duet_colPCR_rev	ACC CCT CAA GAC CCG TTT AG

Table S4. Nanobody TIR sequences identified in this study.

TIR^{orig}	TCGAGG**ATG**GCTCAG
$TIR^{SynEvo1}$	TGGTAA**ATG**GCCCAA
$TIR^{SynEvo2}$	GAATAT**ATG**GCTCAA

References

1. Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., and Endy, D. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. Nat. Methods 10, 354–360.
2. Borujeni, A. E., Channarasappa, A. S., and Salis, H. M. (2013) Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. Nucleic Acids Res. 42, 2646–2659.
3. Woo, S., Yang, J., Kim, I., Yang, J., Eun, B., Kim, S., and Yeol, G. (2013) Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. Metab. Eng. 15, 67–74.
4. Nørholm, M. H. H. (2010) A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. BMC Biotechnol. 10:21.
5. Silva-Rocha, R., Martinez-Garcia, E., Calles, B., Chavarria, M., Arce-Rodriguez, a., de las Heras, a., Paez-Espino, a. D., Durante-Rodriguez, G., Kim, J., Nikel, P. I., Platero, R., and de Lorenzo, V. (2013) The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. Nucleic Acids Res. 41, 666–675.
6. Mirzadeh, K., Martínez, V., Toddo, S., Guntur, S., Herrgård, M. J., Elofsson, A., Nørholm, M. H. H., and Daley, D. O. (2015) Enhanced Protein Production in Escherichia coli by Optimization of Cloning Scars at the Vector–Coding Sequence Junction. ACS Synth. Biol. 4, 959–965.
7. Nørholm, M. H. H., Toddo, S., Virkki, M. T. I., Light, S., von Heijne, G., and Daley, D. O. (2013) Improved production of membrane proteins in Escherichia coli by selective codon substitutions. FEBS Lett. 587, 2352–2358.

Paper 2

Selection of highly expressed gene variants in *Escherichia coli* using translationally-coupled antibiotic selection markers

Rennig M.[1], Daley D.O.[2], Nørholm M.H.H.[1]

[1]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Denmark; [2]Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, Sweden.

Jensen M., Keasling J. (eds)
Synthetic Metabolic Pathways.
Methods in Molecular Biology 1671, pp 259-268. Humana Press, New York, NY

Summary

Strategies to select highly expressed variants of a protein coding sequence are usually based on trial and error approaches, which are time-consuming and expensive. We address this problem using translationally-coupled antibiotic resistance markers. The system requires that the target gene can be fused at the 3'-end with a translational coupling element and an antibiotic resistance gene. Highly expressed target genes can then be selected using a fast and simple whole cell survival assay in the presence of high antibiotic concentrations. Herein we show that the system can be used to select highly expressing clones from libraries sampling translation initiation sites.

Keywords: *gene expression, protein production optimization, selection, library screening, antibiotic resistance, translational coupling*

1. Introduction

Bacterial production of recombinant proteins is highly important in the construction of cell factories, and for basic studies on the function, interactions and structure of proteins. Unfortunately, yields are often low, particularly in heterologous hosts, and optimisation of the coding sequence is necessary. Coding sequence optimisations frequently utilise randomised libraries combined with screening approaches to pick out the coding sequence that expresses to the highest level. Whilst the randomised libraries are simple and inexpensive to make, the screening steps are usually costly and time-consuming[1].

Translational coupling is a natural phenomenon in bacteria, where initiation of translation is dependent on the successful translation of an upstream sequence, e.g. in the tryptophan operon of *E. coli* between the *trpB* and *trpA* genes[2] or in the tightly controlled stoichiometric expression of genes in the ATP operon encoding the subunits of the ATP synthase complex[3]. This mechanism has been exploited to synthetically connect the translation of a target gene to a reporter gene without creating protein fusions[4]. In this elegant study, Mendez-Perez and co-workers designed synthetic translational-coupling devices using the knowledge that mRNA secondary structure can mask a Shine-Dalgarno (SD) site, but may re-fold into a less inhibitory structure by the action of an upstream translating ribosome[4]. Applying similar design principles, we created an extended toolbox of sequences that couple with different efficiency and to different antibiotic resistance markers (unpublished data). Those reporters, when coupled to the production of a protein of interest, offer specific growth advantages to the host organism and, therefore, represent a highly

attractive alternative to e.g. fluorescent proteins, enabling screening of very large libraries[5-7]. The extended toolbox of coupling devices enables the selection of gene expression variants from virtually no expression to g/L industrial scale (unpublished data).

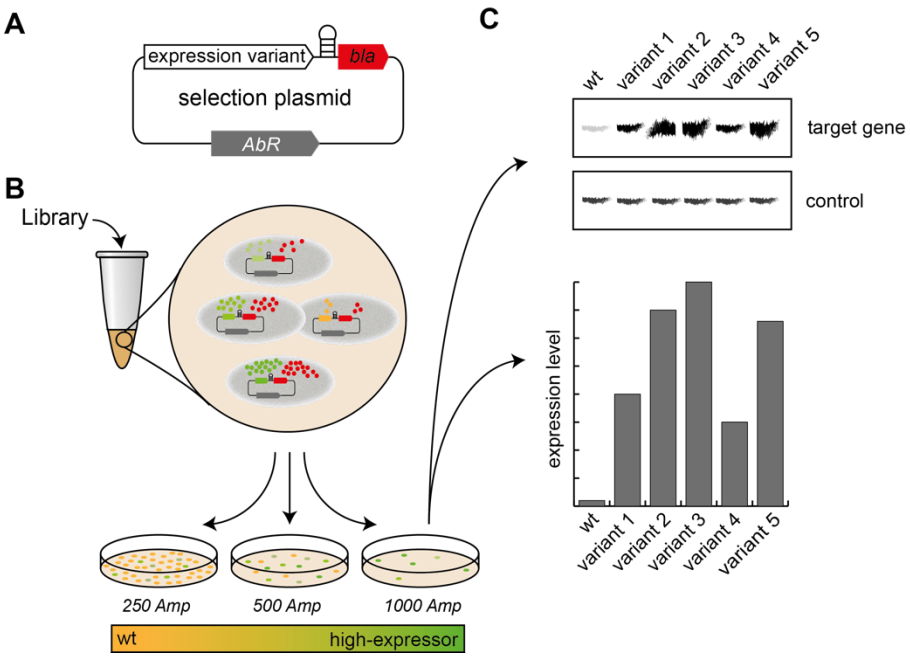


Figure 1. Schematic illustration of protein production optimization with the translationally-coupled antibiotic selection system. (A) Selection plasmid construct with beta-lactamase gene (bla) encoding the ampicillin selection marker. (B) Selection of high-expressing variants by plating on LB agar plates containing 0.25 mg/mL, 0.5 mg/mL and 1 mg/mL ampicillin (Amp). (C) Western blot analysis of the selected optimized variants, on 1 mg/ml ampicillin and expression level estimation of the high-expressing variants.

In this chapter we present our protocols for plasmid construction and selection of high-expressing clones using these translationally-coupled antibiotic resistance markers. An overview of the method is presented in Figure 1. In the first three days the selection plasmid is constructed and isolated using uracil excision DNA assembly (Figure 1A). The plasmid will contain the coding sequence to be expressed (termed expression variant) fused to a region encoding the translational-coupling device (depicted as an mRNA hairpin) and finally, a region encoding for the antibiotic resistance marker (shown is the coding region for ß-lactamase). In the following four days a plasmid library is constructed and each plasmid variant is tested by antibiotic sensitivity (Figure 1B). The library is constructed in a simple PCR reaction using degenerate primers that randomise either the Translation Initiation Region (TIR) or the coding sequence[8,11]. The entire library is isolated and transformed into a standard *E. coli* expression strain and gene expression is induced in liquid cultures. Bacteria are plated on LB agar plates containing different concentrations of the

selectable antibiotic and high-expressing clones can be selected by their ability to withstand the antibiotics. When we applied the protocol to a plasmid encoding a secreted single-chain antibody fragment we were able select a number of gene variants that expressed to higher levels than the wild type coding sequence (Figure 1C). These gene variants differed only in the TIR[8]. Shown are Western blots decorated with the corresponding antibody anti-sera (top panel) and a loading control anti-sera (bottom panel). The protocol enables the selection of the high-expressing clones in one week starting from initial cloning and using an absolute minimum of resources and only standard molecular biology equipment.

2. Materials

2.1 Components for plasmid construction

1. A set of oligonucleotides to integrate the translational coupling element and the antibiotic resistance gene into the desired expression plasmid (see **Note 1** for design).
2. PfuX7 DNA polymerase with 10X reaction buffer (see **Note 2**).
3. dNTP mix (10 mM each), sterile H2O and MgCl2 (final concentration 50 mM) (see **Note 3**).
4. 0.2 mL PCR tubes (VWR).
5. Thermocycler (BioRad).
6. DpnI restriction enzyme (New England Biolabs) (see **Note 4**).
7. 1x TAE buffer. Prepare from 50x stock solution (Life Technologies) with H2O. Can be stored at room temperature.
8. SeaKem® LE Agarose (Lonza).
9. RedSafe Nucleic Acid Staining Solution (iNtRON Biotechnology).
10. NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel) or any other PCR Clean-up Kit.
11. NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific).
12. 100 ng of each PCR fragment, USER enzyme (New England Biolabs) and provided reaction buffer (see **Note 5**).

2.2 Components for transformation and plasmid propagation

1. *Escherichia coli* strain NEB5α for cloning. Chemically competent cells of NEB5α are obtained as described elsewhere (9).
2. Bacteria are cultivated in Luria-Bertani broth (20 g/L in H2O) (Sigma Aldrich) and plated on Luria-Bertani agar (35 g/L in H2O) (Sigma Aldrich).

3. Antibiotics: 100 μg/mL ampicillin, 50 μg/mL kanamycin, 34 μg/mL chloramphenicol, 50 μg/mL spectinomycin and 10 μg/mL gentamycin (see **Note 6**).
4. 1.5 mL microfuge tubes (Eppendorf).
5. Thermomixer (Eppendorf).
6. 50 mL reaction tubes (Sarstedt).
7. Shaking incubator at 37 °C for 50 ml reaction tubes.
8. Plate incubator at 37 °C.
9. QIAprep Spin Miniprep Kit (Qiagen) or any other plasmid DNA purification Kit.

2.3 Components for expression and selection

1. For library expression and selection we use the chemically competent *E. coli* strain BL21(DE3) pLysS (Novagen) (see **Note 7**).
2. Bacteria are cultivated in Luria-Bertani broth (20 g/L in H2O) (Sigma Aldrich) and plated on Luria-Bertani agar (35 g/L in H2O) (Sigma Aldrich).
3. 1M Isopropyl β-D-1-thiogalactopyranoside (IPTG) stock solution in H2O (see **Note 8**).
4. Antibiotics for selection: ampicillin, kanamycin, chloramphenicol or spectinomycin (see **Note 9**).
5. 50 mL reaction tubes (Sarstedt).
6. Shaking incubator at 37 °C for 50 mL reaction tubes.
7. Plate incubator at 37 °C.

3. Methods

3.1 Primer design and selection set-up

1. Before constructing the selection plasmid, the expression level of the target gene needs to be evaluated carefully. We have developed a set of different translational coupling devices that vary in their coupling efficiency (Figure 2). For low expressed target genes we recommend the strong coupling device. For genes already expressing to decent levels, a weaker coupling device is advisable. This way the use of very high antibiotic concentrations for selection is avoided and the dynamic range of the selection system is better exploited (see **Note 10**). The coupling device is introduced with PCR oligonucleotides - the major part with a "forward" oligonucleotide for amplification of the antibiotic resistance gene (here denoted AbR gene fwd). The reverse oligonucleotide for amplifying the target gene introduces the second part of the coupling device (here denoted target gene rev.) (Figure 3).

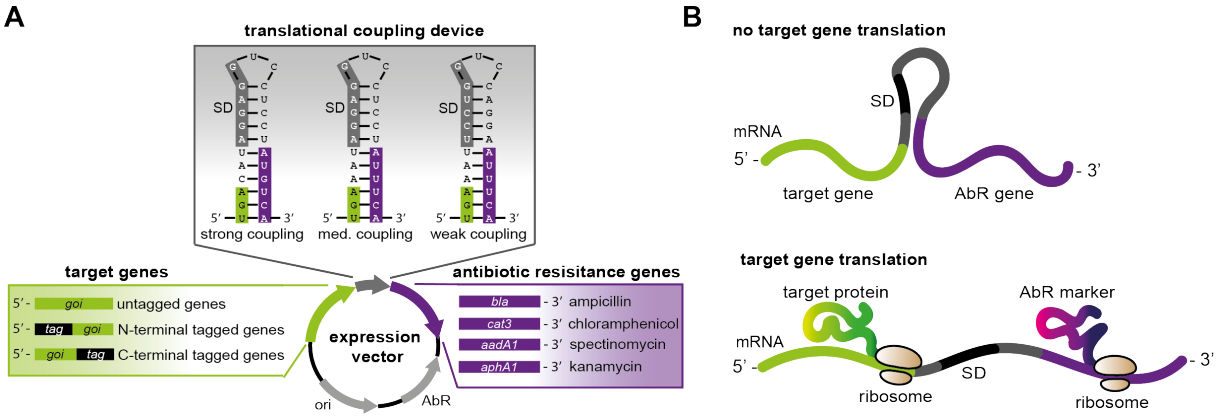


Figure 2. Different modules can be combined for optimal antibiotic selection via translational coupling. (A) The antibiotic selection system optimizes the expression of untagged, N-terminal tagged or C-terminal tagged genes. The target gene (green box) is combined with one out of three translational coupling devices (grey box). The three different hairpins vary in their coupling efficiency due to altered start codons (shaded purple regions) and Shine-Dalgarno (SD) sequences (shaded grey regions). Four different antibiotic resistance genes (AbR, purple box) can be used as reporters. (B) Illustration of the principle of translational coupling; if the mRNA from the target gene (green) is not translated, the ribosome-binding site (black) for the reporter gene (purple, AbR gene) is not accessible. If the target gene is translated, the ribosome will melt the hairpin structure and enable a ribosome to bind to the SD sequence of the reporter gene mRNA and start translation.

2. The target gene can either be amplified together with the backbone or can be cloned into a new backbone. When amplified with the backbone primer set 1 is not needed (Figure 3). In case that primer set 1 is needed, the design will be similar to primer set 3. The reverse primer for amplifying the target gene introduces parts of the coupling device (target gene rev.) and the only necessary modification to the gene of interest ensures that TGA is used as a stop codon. Moreover, the reverse primer for amplifying the target gene should ideally anneal with a melting temperature of ca. 60 °C to the sequence upstream from the stop codon. Typically, this takes roughly 20 nucleotides for the anneal part resulting in a total oligonucleotide size of ca. 36.
3. The major part of the translational coupling device is introduced with the forward primer for amplification of the antibiotic resistance gene (AbR gene fwd.). Besides encoding the device, the primer should anneal with a melting temperature if ca. 60 °C to sequence downstream from the start codon. This will typically result in oligonucleotide size of ca. 40 nucleotides. Note that the first amino acid after the start codon is still part of the coupling device and will always be TCA (encoding for serine). The ampicillin and kanamycin resistance genes, bla and aphA1, already encode serine as the second amino acid. For the spectinomycin and

chloramphenicol resistance genes, aadA1 and cat, the addition of serine in this position does not influence functionality (see **Note 11**).

4. Following the PCR, cohesive ends with a melting temperature of ca. 23 °C are formed by uracil excision (see **Note 12**) and the resulting nicked circular DNA can be transformed into chemically competent *E. coli*. Details about uracil excision DNA assembly can be found elsewhere (10).

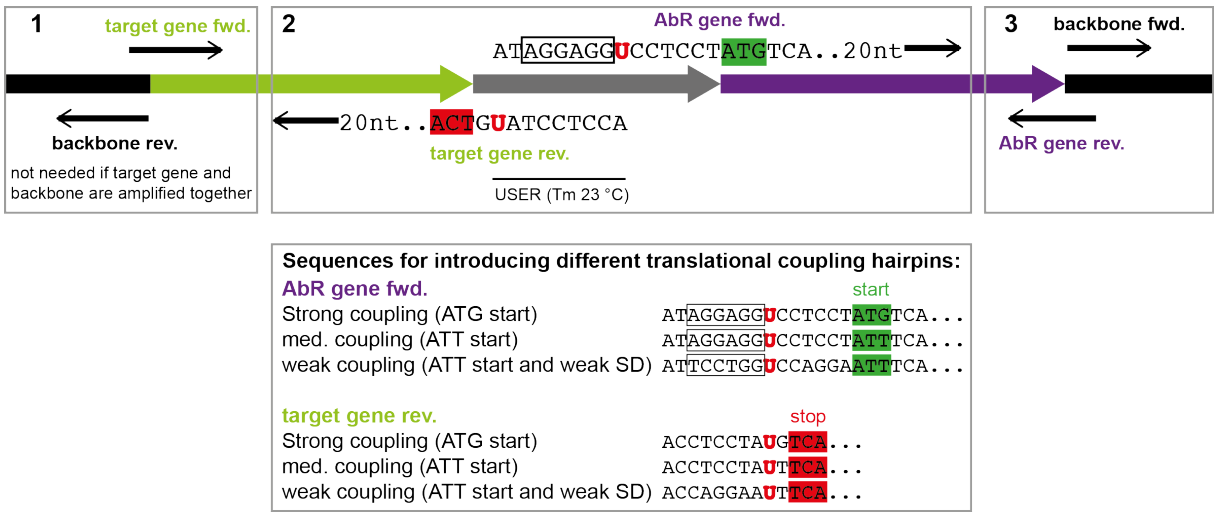


Figure 3. Illustration of the design of oligonucleotides that introduce a translational coupling device and an antibiotic resistance gene for selection. Primer pairs denoted in boxes 1, 2 and 3 have complementary 5′ ends as exemplified with sequence detail in box 2.

3.2 Plasmid construction and propagation

1. Mix all reagents for a 50 µL PCR reaction in a 0.2 ml PCR tube: 5 µL of 10X reaction buffer, 5 µL of forward primer, 5 µL of reverse primer, 2 µL dNTP mix (40 mM), 1.2 µL 2M MgCl2, 1 µL PfuX7 DNA polymerase, 100 ng of template DNA, fill up to 50 µL with sterile H2O.
2. Amplify the different fragments in a Thermocycler using a Touchdown-PCR program (95 °C for 5 min, 10 cycles of 95 °C for 45 sec, 55-65 °C for 45 sec (increment 1 °C in each cycle), 72 °C for 5 min, followed by 20 cycles of 95 °C for 45 sec, 55 °C for 45 sec, 72 °C for 5 min and one final elongation step for 8 min at 72 °C (see **Note 13**).
3. Prepare a 1% agarose gel by heating up 1 g of agarose in 100 ml of 1 x TAE buffer. Add DNA stain. Analyze 5 µlLof the PCR product by agarose gel electrophoresis (e.g. 90 V, 30 min).
4. Add 2 µL DpnI to the reaction mix to digest the original plasmid at 37 °C for at least 1 h. Inactivate DpnI for 5 min at 80 °C.

5. Purify the PCR fragments using a PCR Clean-up Kit, following the manufacturer's instructions.
6. Mix 100 ng of each fragment. Add 1 µL of USER enzyme and 1 µL of buffer. Fill up with sterile H2O to a final volume of 10 µL.
7. Start the USER reaction in a Thermocycler: 37 °C for 15 min, 23 °C for 15 min, 10 °C for 15 min (see **Note 14**).
8. Mix the whole USER reaction with 100 µL of chemically competent *E. coli* cells. Use a cloning strain, such as NEB5α.
9. Incubate on ice for 30 min.
10. Heat shock for 1 min at 42 °C.
11. Incubate on ice for 3 min.
12. Add 0.9 mL of LB broth without antibiotics for recovery and incubate at 37 °C with shaking for 1 h.
13. Plate 0.1 mL (10%) of the transformation mix on a LB agar plate containing the appropriate antibiotics.
14. Harvest the rest of the transformation mix by centrifugation at 4000 x g in a tabletop centrifuge and plate on an LB agar plate containing the appropriate antibiotics.
15. Incubate for 16 h at 37 °C.
16. Inoculate 5 mL LB media containing the appropriate antibiotics (in a 50 ml reaction tube) with a single colony from the transformation plates and incubate for 16 h at 37 °C with shaking.
17. Harvest the cells at 6500 x g and purify the plasmid using a plasmid DNA purification Kit (e.g. QIAprep Spin Miniprep Kit), follow the manufacturer's instructions.
18. Prepare an expression library (see **Note 15**).

3.3 Expression and selection

1. Transform the library and the wild type plasmid into an expression strain, e.g. *E. coli* BL21(DE3) pLysS. To do so, follow steps 7-11 in section 3.2. Transform ca. 5 µL of library into 50 µL of commercially competent cells, and 1 µL of wild type plasmid into 15 µL of commercially competent cells. Heat shock for 45 sec at 42 °C.
2. After 1 h of recovery transfer the transformation mix to a 50 mL reaction tube containing 5 ml of LB media with appropriate antibiotics and incubate for 16 h at 37 °C with shaking.
3. Prepare plates for selection. Use LB agar containing the inducing agent, we use IPTG in a final concentration of 1 mM, and different concentrations of the selective antibiotic (see **Note 16**).

We recommend preparing selection plates with at least 5 different antibiotic concentrations. Remember to prepare two plates of each concentration, one for the library and one for the wild type.

4. After 16 h, start an expression culture by inoculating 5 mL fresh LB medium containing appropriate antibiotics with 100 µL of over night culture (1:50). Incubate at 37 °C with shaking.
5. Grow the culture to an OD600 of 0.3 – 0.5 (ca. 2 h) and induce expression with IPTG (final concentration 1 mM) of any other agent needed to induce expression. Incubate at 37 °C with shaking.
6. After 2 h of expression, plate 0.2 ODU (ca. 100 µL) on each selection plate. Incubate for up to 40 h at 37 °C. On antibiotic concentrations where no growth can be seen for the wild type, high expressing variants can be selected.
7. Select variants and check with sequencing (see **Note 17**).
8. Re-transform the selected variants to confirm the selection of a highly translated variant. Follow steps 1-5 (skip step 3) and check for expression levels, e.g. by western blot analysis.

4. Notes

1. The forward primer of the target gene and the reverse primer of the antibiotic resistance gene form the translational coupling device. Oligonucleotides contain an incorporated uracil for USER cloning.
2. Alternatively, Phusion U Hot Start DNA Polymerase (Thermo Scientific) can be used.
3. If Phusion U Hot Start DNA Polymerase is used, 1.5 µL of 1 M DMSO is added to a 50 µL reaction.
4. 2 µl FastDigest DpnI (Thermo Scientific) were added directly to a 50 µL PCR mix after PCR reaction, since DpnI works effectively in the reaction buffers used for PCR.
5. We recommend a ligation step after the USER reaction, if electro-competent cells are used. In this case, we recommend using T4 ligase buffer for the USER reaction.
6. Antibiotic concentrations indicated are those that should be used for the backbone antibiotic resistance. Make sure that the backbone resistance and the resistance used for selection are different.
7. Be aware that when using a plasmid with chloramphenicol resistance, either in the backbone or as selection module, BL21(DE3) without pLysS needs to be used, as the pLysS plasmid (and derivatives) confer resistance to chloramphenicol.

8. The type of inducing agent depends on the promoter used. We are using a (DE3) T7 promoter and therefore typically use IPTG for induction.
9. Concentrations depend on the target gene expression level and the chosen coupling efficiency.
10. The coupling device can be changed easily in a one-PCR step followed by a one fragment USER reaction.
11. Note that the chosen resistance gene for selection should not be present as a backbone resistance gene on any other plasmid transformed. We obtained the best results when using the ampicillin resistance gene for selection.
12. We recommend choosing a USER cloning overlap in the backbone sequence, as it facilitates the exchange of the antibiotic resistance gene, if necessary.
13. Note that the extension time depends on the size of the PCR product and the processivity of the DNA polymerase that is used.
14. When using electro-competent cells, add 1.5 μL of T4 ligase and incubate for 15 min at room temperature prior to transformation.
15. Library construction is not part of this protocol. Sequence variation may be sampled throughout the gene coding sequence or expression vector. We construct libraries based on the protocol "Codon optimizing for increased membrane protein production: A minimalist approach" (11). Additionally, we expand the library by also randomizing the six nucleotides upstream of the start codon (8).
16. Make sure that the LB agar has cooled down to at 60 °C before adding the inducing agent and antibiotics.
17. Sequencing after selection is of outmost importance. It can occur that the cell starts mutating in ways to favour the expression and usage of the antibiotic resistance gene and a false positive is selected.

5. References

1. G.L. Rosano and E.A. Ceccarelli (2014) Recombinant protein expression in Escherichia coli: advances and challenges, Frontiers in Microbiology. 5, 1–17.
2. S. Aksoy, C.L. Squires, and C. Squires (1984) Translational Coupling of the trpB and trpA Genes in the Escherichia coli Tryptophan Operon, Journal of Bacteriology. 157, 363–367.
3. G. Rex, B. Surin, G. Besse, et al. (1994) The Mechanism of Translational Coupling in Escherichia coli, 269, 18118–18127.
4. D. Mendez-Perez, S. Gunasekaran, V.J. Orler, et al. (2012) A translation-coupling DNA cassette for monitoring protein translation in Escherichia coli, Metabolic Engineering. 14, 298–305.
5. M. Rennig, V. Martinez, D.O. Daley, et al. Synthetic evolution of the translational initiation region maximizes protein production in bacteria, In preperation.

6. E. Massey-Gendel, A. Zhao, G. Boulting, et al. (2009) Genetic selection system for improving recombinant membrane protein expression in E. coli, Protein Science. 18, 372–383.
7. N. Gul, D.M. Linares, F.Y. Ho, et al. (2014) Evolved Escherichia coli strains for amplified, functional expression of membrane proteins, Journal of Molecular Biology. 426, 136–149.
8. R. Tan, X. Jiang, A. Jackson, et al. (2003) E. coli selection of human genes encoding secreted and membrane proteins based on cDNA fusions to a leaderless β-lactamase reporter, Genome Research. 13, 1938–1943.
9. K. Mirzadeh, V. Martínez, S. Toddo, et al. (2015) Enhanced Protein Production in Escherichia coli by Optimization of Cloning Scars at the Vector–Coding Sequence Junction, ACS Synthetic Biology. 4, 959–965.
10. H. Inoue, H. Nojima, and H. Okayama (1990) High efficiency transformation of Escherichia coli with plasmids, Gene. 96, 23–28.
11. A.M. Cavaleiro, S.H. Kim, S. Seppälä, et al. (2015) Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning, ACS Synthetic Biology. 4, 1042–1046.
12. K. Mirzadeh, S. Toddo, M.H.H. Nørholm, et al. (2016) Codon Optimizing for Increased Membrane Protein Production: A Minimalist Approach, In: Heterologous Expression of Membrane Proteins: Methods and Protocols, Methods in Molecular Biology, pp. 53–61, Springer Science+Business Media New York.

Paper 3

A gram positive platform for selection of highly expressed gene variants

Roberto Ferro[1,2]*, Maja Rennig[1]*, Cristina Hernandez Rollan[1], Daniel O. Daley[3,4] and Morten H. H. Nørholm[1,4]

[1]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[2]Department of Plant and Environmental Science, University of Copenhagen, 1871 Frederiksberg, Denmark
[3]Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, Sweden.
[4]CloneOpt AB, Upplands Väsby, Sweden

* *Contributed equally*

Submitted to Microbial Cell Factories

Abstract

Background

The market for recombinant proteins is on the rise, and Gram-positive strains are widely exploited for this purpose. *Bacillus subtilis* is a profitable host for protein production thanks to its ability to secrete large amounts of proteins, and *Lactococcus lactis* is an attractive production organism with a long history in food fermentation.

Results

We have developed a synbio approach for increasing gene expression in two Gram-positive bacteria. First of all, the gene of interest was coupled to an antibiotic resistance gene to create a growth-based selection system. We then randomised the translation initiation region (TIR) preceding the gene of interest and selected clones that produced high protein titres, as judged by their ability to survive on high concentrations of antibiotic. Using this approach, we were able to significantly increase production of two industrially relevant proteins; sialidase in *B. subtilis* and tyrosine ammonia lyase in *L. lactis*.

Conclusion

Gram-positive bacteria are widely used to produce industrial enzymes. High titres are necessary to make the production economically feasible. The synbio approach presented here is a simple and inexpensive way to increase protein titres, which can be carried out in any laboratory within a few days. It could also be implemented as a tool for applications beyond TIR libraries, such as screening of synthetic, homologous or domain-shuffled genes.

Background

The advent of recombinant protein technology has enabled commercial applications for biopharmaceutical proteins and industrial biocatalysts not possible when the only option was to extract proteins from their original hosts. The enzyme market in particular has bloomed as a result of the production costs approaching those of the chemical industry[1]. To achieve low costs, enzymes are produced in large quantities by exploiting cell factories and large-scale bioreactors, but continuous improvements in the design of cell factories are needed to keep the production competitive and open markets for new products[2].

Various rational engineering approaches for cell factories are routinely employed to improve production. In the initial process a suitable expression host needs to be selected. Despite the documented role of *Escherichia coli* in molecular biology, other bacterial expression systems have been explored in biotechnology, taking advantage of divergent metabolism, secretion capability and biosafety of their protein-based products[3,4]. Amongst them, various Gram-positive bacteria are of great interest due to e.g. their highly efficient protein secretion and GRAS status. *Bacillus subtilis* is routinely used industrially for its ability to secrete large amount of enzymes, which simplifies protein recovery and purification, leading to yields up to 20-25 g/L[5]. *Lactococcus lactis* has a long history of use in food microbiology and in the dairy industry[6], and it has lately risen as an emerging alternative for production of membrane proteins[7,8], secreted proteins[9,10] and plant-based proteins and secondary metabolites[11-13].

Previously we have shown that protein production in Gram-negative bacteria can be significantly increased by synthetically evolving a part of the translation initiation region (TIR)[14]. In expression clones the TIR extends from the region upstream of the Shine-Dalgarno (SD) sequence to the 5th or 6th codon of the gene of interest[15]. Whilst most TIRs function, they can support higher production titres if they are evolved with an appropriate selection pressure. For example, the TIR can be optimised by creating large libraries of randomised TIRs and selecting one that produces the most protein[16,17].

Yet screening approaches for those libraries are limited. High throughput screening methods like fluorescence activated cell sorting (FACS) or droplet microfluidics enable the assessment of large libraries[18-20] but their use is restricted to phenotypes associated with a fluorophore to effectively screen for e.g. increased protein production[21]. Fusions with reporter proteins, such as fluorescent proteins, can also compromise the expression, solubility and bioactivity of a protein and are not suitable in industrial set-ups[22,23]. The availability of other types of biosensors is limited and developing one for a new target is a laborious and time consuming process[24].

For this purpose, we have established a phenotypic screening approach for screening TIR libraries in *E. coli*[25]. The approach utilises an antibiotic resistance gene that is translationally coupled to the gene of interest. In the bicistronic mRNA design, a hairpin-like structure separates an upstream gene of interest from a downstream antibiotic selection marker, thereby sequestering the SD site of the latter. Only upon efficient translation of the upstream gene, antibiotic resistance is obtained

because the helicase activity of the translating ribosome allows expression of the downstream resistance gene (Figure 1A). In this study we set out to determine if the same synbio approach for optimising and selecting TIRs, would lead to increased production levels in industrially relevant Gram-positive bacteria.

Results

Characterization of a translational coupling device in Gram-positive hosts

We first set out to explore the applicability of a translational coupling design that was previously developed in *E. coli*[25]. To this end we constructed a plasmid for *L. lactis* and *B. subtilis* that contained the gene encoding green fluorescent protein (*gfp*), coupled by a sequence with hairpin-forming (hp) propensity to a chloramphenicol resistance gene (*gfp*-hp-*CmR*) (Figure 1A).

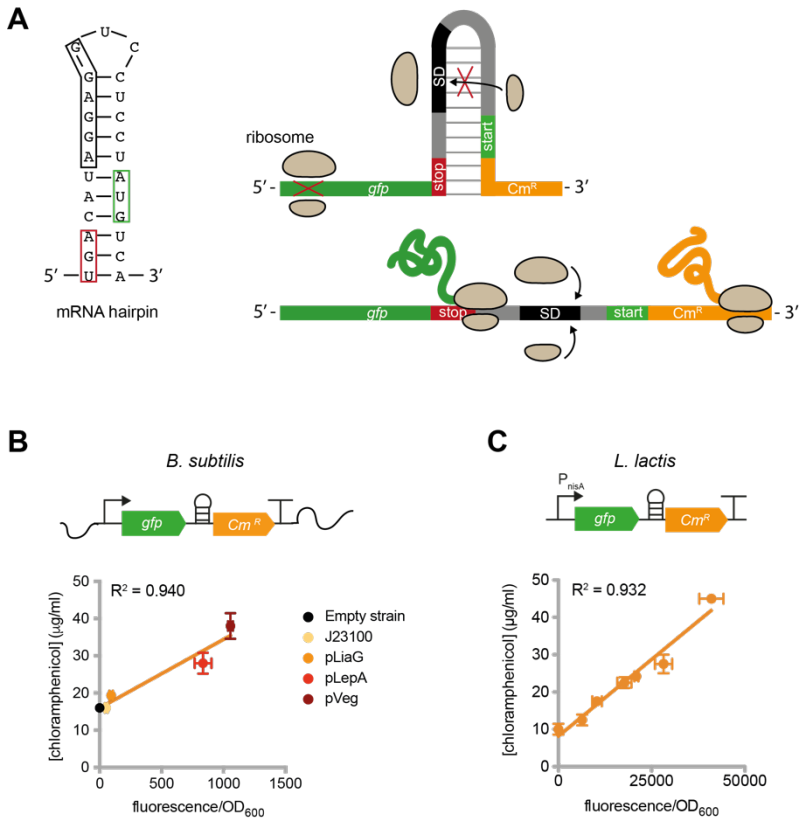


Figure 1: Testing of a translational coupling device in *Bacillus subtilis* and *Lactococcus lactis*. (A) To asses the efficacy of a specific translational coupling device in *B. subtilis* and *L. lactis* an mRNA hairpin structure was sandwiched between the gene encoding green fluorescent protein (gfp, green) and the chloramphenicol resistance gene (CmR, orange). Left side: Shown is the predicted structure of the translational coupling device. Presumably, the stem of the mRNA hairpin structure consists of 11 nucleotide pairs comprising the stop codon of the upstream gene (red box) and the start codon of the downstream gene (green box). The ribosome binding site (black box) of the downstream gene is designed to be masked by the secondary mRNA structure. Upper right side: When the upstream gene (green, gfp) is not translated, the mRNA hairpin structure will not be resolved and the ribosome

binding site of the downstream gene (orange, CmR) remains inaccessible for the ribosome. Therefore, there is no translation of the downstream gene. Lower right side: When a ribosome translates gfp, the ribosome's helicase activity will melt the secondary mRNA structure which makes the ribosome binding site accessible and the chloramphenicol resistance gene can be translated. The correlation between protein production, determined as fluorescence normalized for cell density, and chloramphenicol resistance, determined as minimal inhibitory concentration (MIC), was determined for genome-based expression in *B. subtilis* (B) and for plasmid-based expression in *L. lactis* (C).

The *gfp*-hp-*CmR* construct was expressed from the nisin-inducible promoter P_{nisA} on a pNZ8048 vector in *L. lactis*[26]. In *B. subtilis* a set of four different constitutive promoters of increasing strength (P_{J23101}, P_{liaG}, P_{lepA} and P_{veg}[27]) were used and the constructs were integrated into the *amyE* locus on the chromosome[28]. *L. lactis* cultures were grown overnight, diluted in the morning and induced with variable concentrations of nisin (0.25-10 ng/mL); *B. subtilis* cultures were treated the same way, but did not require induction. Fluorescence was measured when the cultures reached late exponential phase and was normalized by cell density. At the same time resistance was assessed by plating on different concentrations of chloramphenicol. In both *L. lactis* and *B. subtilis* fluorescence levels showed a linear correlation ($R^2 > 0.9$) with resistance to chloramphenicol, the latter measured as the minimum inhibitory concentration (MIC) (Figure 1B, C). This demonstrates that the translational coupling device in combination with a chloramphenicol resistance gene can work over a broad range as a reporter of gene expression in these two Gram-positive model bacteria.

Characterization of translation initiation region libraries in L. lactis and B. subtilis

Next we introduced sequence diversity in a specific part of the translation initiation region (TIR) – an approach that has proven successful for expression optimization in *E. coli*[14]. In this experiment the six nucleotides upstream of the start codon are randomized and the second and third codon of the open reading frame are concurrently substituted with synonymous codons (Figure 2A). Depending on the nature of the 2nd and 3rd amino acid in the protein sequence, this system generates a maximum library size of about 150,000 variants and up to 1000-fold variation in gene expression levels[14].

In the case of *gfp*, a library with about 50,000 variants was constructed in both hosts. In *B. subtilis* we chose a strong constitutive promoter (P_{veg}) to drive expression of *gfp*. The library was integrated into the *amyE* locus on the chromosome using an integrative vector propagated in *E. coli*. In *L. lactis* we used the pNZ8048 plasmid with the inducible nisin promoter controlling the expression of *gfp*.

Both libraries were based on the constructs used for the initial characterization of the translational coupling device *gfp*-hp-*CmR*. Variability was introduced by PCR using degenerate oligonucleotides that randomized the TIRs. The TIR region of five arbitrary clones from each library were sequenced to validate diversity in the libraries. A total of 96 clones were randomly picked and cultivated. Cultures were back-diluted after overnight incubation, induced when required and expression levels were assayed after five hours. The fluorescence levels varied greatly amongst the clones with the highest and the lowest gfp-expressing clones differing by up to 1000-fold (Figure 2B and C). This observation confirms that this part of the TIR is an important determinant of expression levels in a broad range of bacterial species.

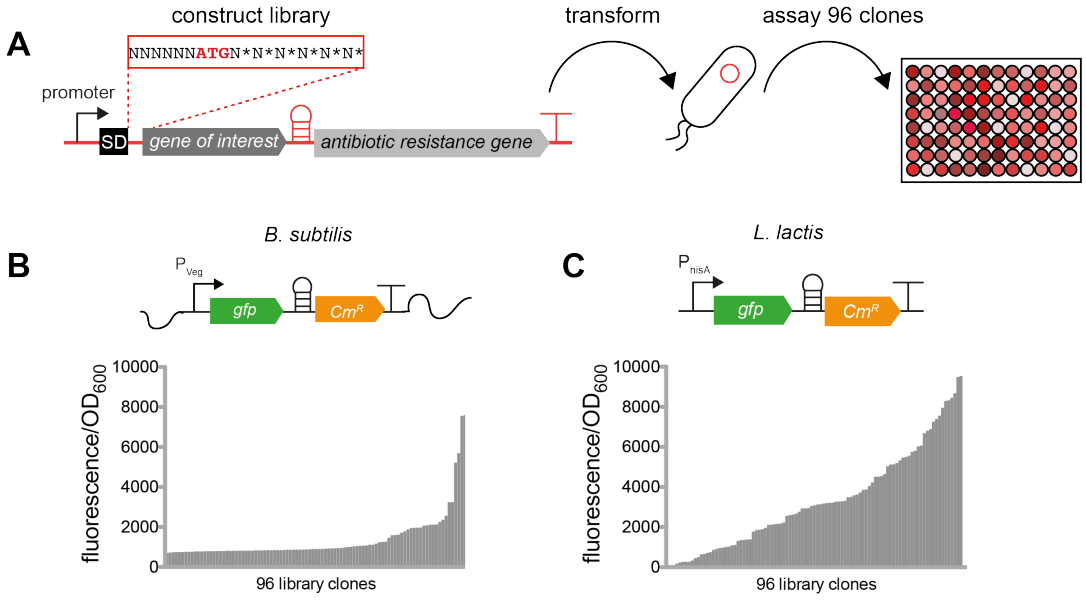


Figure 2: A part of the translation initiation region (TIR) affects expression in *Bacillus subtilis* and *Lactococcus lactis*. (A) TIR libraries (NNNNNNATGN*N*N*N*N*N*, see main text for further details) were constructed by PCR, transformed into the host strains and individual library clones were grown to asses the expression level ranges in *B. subtilis* and *L. lactis*. Fluorescence normalized by cell density was determined for 96 library clones for *B. subtilis* (B) and *L. lactis* (C). The *B. subtilis* library was expressed from the genome whereas the *L. lactis* library was expressed from a plasmid.

Antibiotic-based selection of high-producing variants from TIR libraries

To select for the best TIRs, libraries were plated on solid media with increasing concentrations of chloramphenicol. The amount of colonies appearing on the plates decreased whereas the average fluorescence intensity increased with rising concentrations of antibiotic (Figure 3A). Clones were randomly picked from the plates, recovered and grown overnight without antibiotic selection before measuring the fluorescence in a microplate reader (Figure 3B). This analysis showed that

the likelihood of isolating a highly fluorescent clone increased with increasing concentrations of antibiotics (Figure 3C). In contrast to the plate- based selection setup, we were unable to enrich for high expressing clones in liquid media supplemented with different concentrations of chloramphenicol (data not shown).

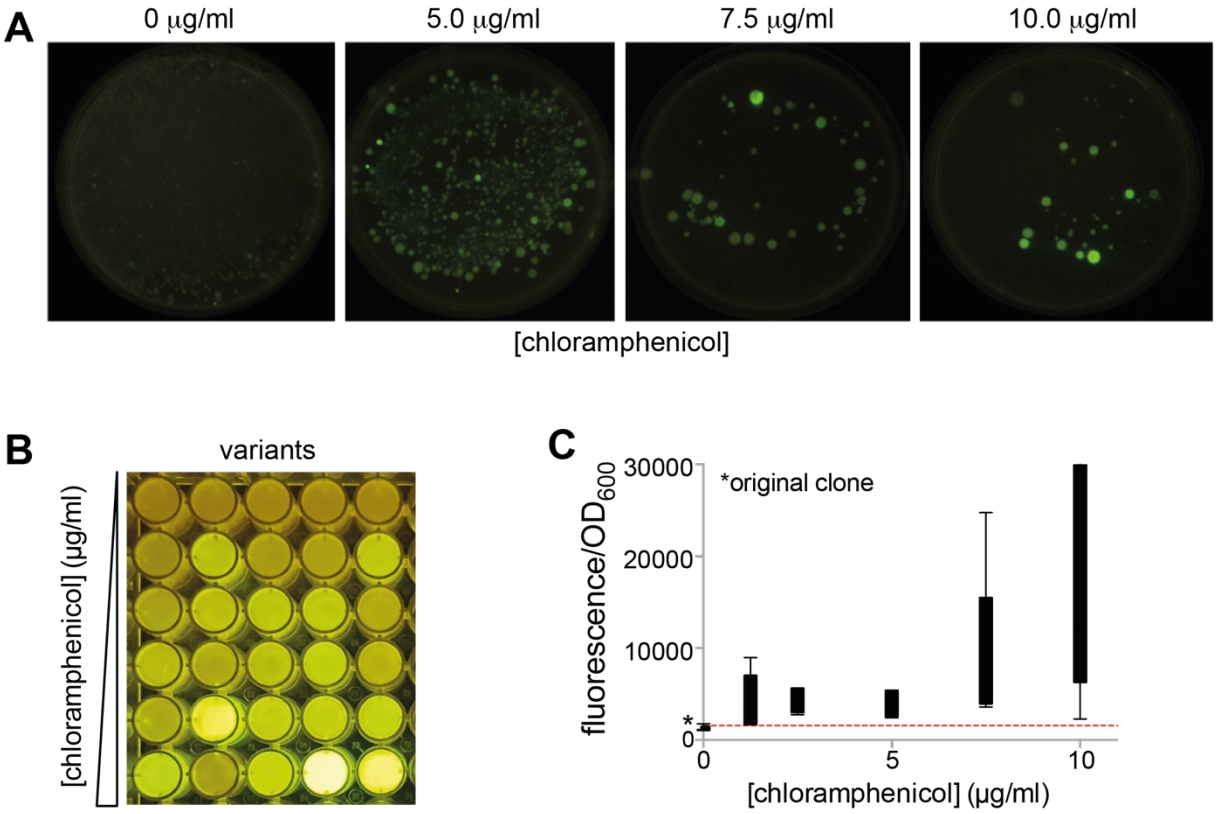


Figure 3: Antibiotic selection of maximal protein production in a Gram-positive bacterium. (A) *B. subtilis* TIR libraries expressing gfp were grown on agar plates with different chloramphenicol concentrations. GFP production levels on the different antibiotic concentrations were assayed by exposing agar plates to long-wave UV light. Five individual colonies were randomly picked from these plates and assayed for expression levels assessed by fluorescence per cell density in a microplate reader (B-C).

Optimization of industrially relevant proteins

Finally, we explored optimization of clones for production of different industrially relevant proteins in the two Gram-positive hosts. For *L. lactis* we chose to optimize expression of a tyrosine ammonia lyase (TAL) from *Flavobacterium johnsoniae*, which is expressed in the cytoplasm and converts tyrosine into p-coumaric acid in a single enzymatic step[29]. For *B. subtilis* we chose to target a *Micromonospora viridifaciens* sialidase (SIA), a hydrolase that cleaves the sialic residues of glycans[30]. In both cases well established expression set-ups were used that already resulted in high expression levels.

We first integrated the selection module hp-CmR downstream of the genes-of-interest into the plasmids used in the previous studies for the corresponding enzyme production[29,30]. In these experiments, cytoplasmic expression of *tal* in *L. lactis* was driven by the inducible P_{nisA} promoter in a pNZ8048 vector; however, we had to exchange the original chloramphenicol cassette of the vector backbone for an erythromycin resistance gene, to be able to utilize the hp-CmR device for selection. Expression of sia in *B. subitilis* was under control of the strong constitutive P_{32} promoter and secreted by the aid of a CGTase signal peptide encoded in the replicative pDP66K plasmid. In both experiments, we refer to the TIR in the original constructs as the TIR^{orig}.

The TIR^{orig} was then randomized by PCR using degenerate oligonucleotides, employing the same strategy as with the *gfp*-hp-*CmR* library. After transformation into the hosts and overnight recovery with the appropriate vector backbone antibiotics in liquid culture, both libraries were back-diluted, induced when required and after five hours plated on solid media with increasing concentrations of chloramphenicol. We then determined the Colony Forming Units (CFUs) of each library at different concentrations of chloramphenicol, and compared it with the TIRorig. In both cases we observed that the library produced more CFUs at higher chloramphenicol concentrations than the original construct (Figure 4A, B). As expected, we also observed a reduction in the amount of colonies appearing on the plates as the antibiotic concentration increased.

Colonies were recovered on 75 μg/mL chloramphenicol for *B. subtilis* and 25 μg/mL chloramphenicol for *L. lactis*. Protein levels were assessed by Western blotting using a His-tag antiserum for sialidase and Strep-tag antiserum for TAL (Figure 4C, D). In both cases, the clones isolated at the highest antibiotic concentration displayed a higher protein production compared to the TIR^{orig}. Production of TAL in *L. lactis* was improved by 8-fold (Figure 4D) and production of sialidase in *B. subtilis* was doubled (Figure 4C) from 1 g/L to approximately 2 g/L (Figure 4E). In both cases, activity of the enzymes originating from the optimized TIR^{opt} variants was not compromised by the high level of production (Supplementary Figure 1A, B).

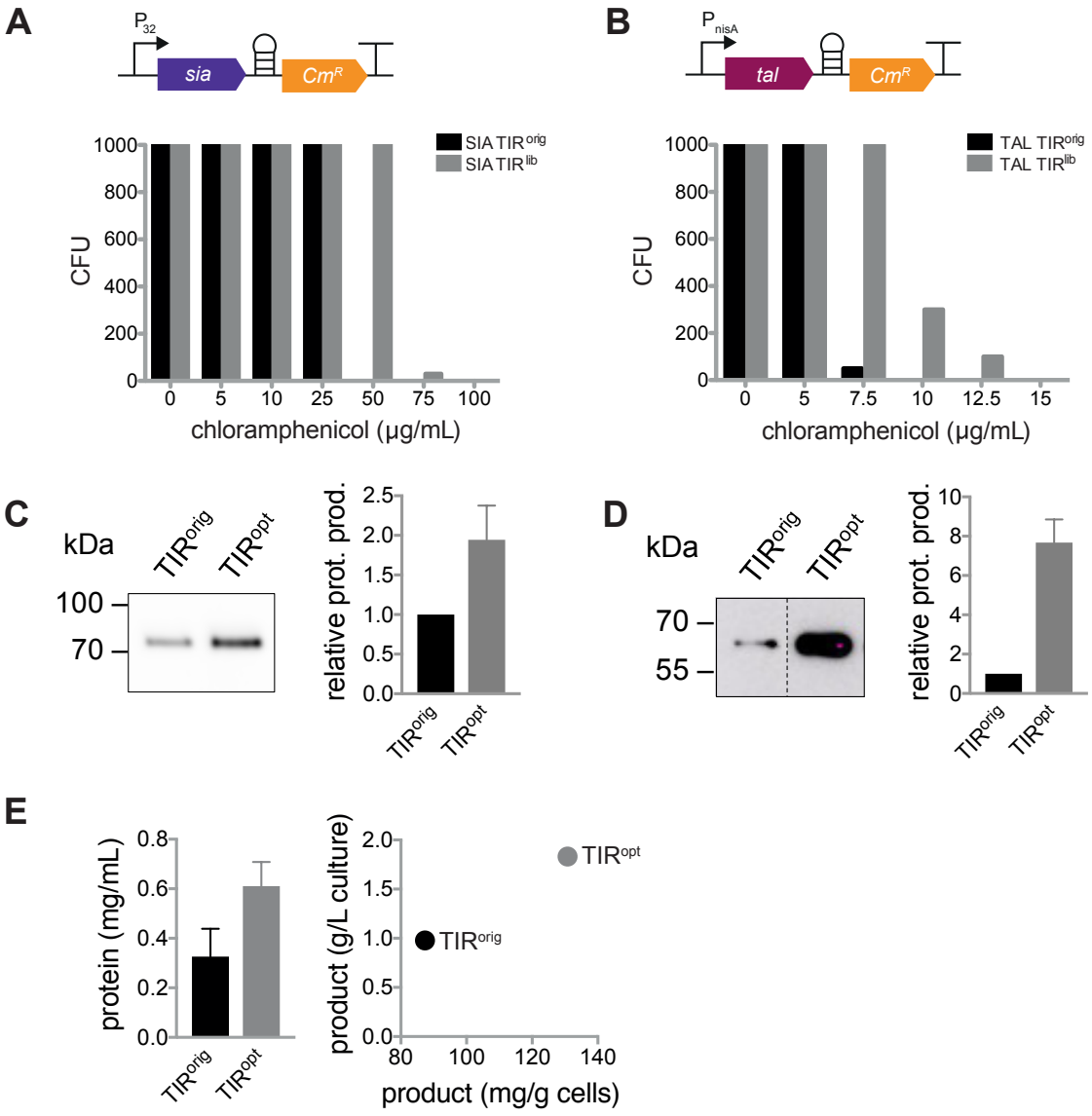


Figure 4: Production optimization of the industrially relevant proteins tyrosine ammonia lyase (TAL) and sialidase (SIA). (A) A His-tagged sialidase-encoding sequence was translationally coupled to the chloramphenicol resistance gene. A TIR library was constructed, transformed into B. subtilis and grown with different chloramphenicol concentrations on agar plates. Colony forming units (CFUs) were counted for all concentrations for the library (TIR^{lib}) and the original, non-randomized clone (TIR^{Orig}) as control. (B) A STREP-tagged tal-encoding sequence was translationally coupled to the chloramphenicol resistance gene. A TIR library was constructed, transformed into *L. lactis* and grown with different chloramphenicol concentrations on agar plates. CFUs were counted for all concentrations for the library (TIR^{lib}) and the original, non-randomized clone (TIR^{orig}) as control. (C) Sialidase production level in the culture supernatant of the best performing clone (TIR^{opt}) was analyzed by Western blot, using an antibody against a His-tag (left panel). The increase in production was analyzed by densitometry and plotted as the relative protein production compared to the original clone (right panel). (D) TAL production level of the best performing clone (TIRopt) was analyzed by Western blotting using an antibody against Strep-tag (left panel). The increase in production was analyzed by densitometry and plotted as the relative protein production compared to the original clone (right panel). (E) Sialidase (TIRopt) was purified via a Ni2+-NTA column and purified product concentration was estimated using a BCA assay (left panel). Product per cell mass and per culture volume was calculated (right panel). The original construct (TIRorig) was used for comparison.

Discussion

A wide range of industrial applications use enzymes as a sustainable alternative to chemical catalysts. These applications include, but are not limited to, the manufacturing of food, paper, detergents and biofuels[31]. However, refinement in production titres rely on enzyme discovery and gene and strain engineering, which are time consuming and labour intensive processes. Existing methods for high throughput screening of gene expression are based on e.g. droplet microfluidics or flow cytometry[18,20]. Despite being extremely efficient and effective, these platforms are expensive to purchase and operate and their availability is limited to large-scale operations. Other screening methods rely on protein fusions, which might alter parameters such as expression, solubility and/or turnover rate[22]. To address this issue, we developed a simple growth based selection system based on translational coupling of an antibiotic resistance gene with the upstream gene sequence that leaves the protein of interest unaltered. Different from previous attempts, our system focuses on Gram-positive bacteria[25], works on genomically integrated targets and does not require tagging of the protein of interest[32]. Nonetheless, tags can be purposefully added to the protein if desired.

We first evaluated the correlation between gene expression and antibiotic resistance. A lack of correlation would be indicative of either poor folding or poor melting of the mRNA hairpin. Instead the correlation coefficient was strong in both hosts, which demonstrates that antibiotic sensitivity is tuneable and that the system responds linearly to all gene expression levels tested. Differences in the dynamic range of screenable chloramphenicol concentrations were observed between the two hosts; presumably the cause for a smaller dynamic range in *B. subtilis* may be due to the initial choice of a poor-performing TIR for *gfp*. The expression was very low, even when driven by a strong promoter. Once we built a combinatorial library randomizing the TIR, the best expressing variants improved in protein production by 8 fold over the variant we used to characterize the mRNA hairpin.

Protein production constructs are often assembled in replicative or integrative plasmids. Previous work has shown that the junction created between the vector and the coding sequence can result in different levels of expression depending on e.g. the restriction site used[14]. The variability is not completely explained by the free energy (ΔG) associated with mRNA folding and seems to be context specific, meaning that the cloning scar will affect the levels of expression in a manner that is not entirely computationally predictable. The creation of a TIR library circumvents the problem

by optimizing the expression of a gene directly in the system in which it will be expressed. Here we demonstrate the effectiveness of removing the cloning scar bias by improving the production of two industrially relevant proteins. We chose to optimize TAL in *L. lactis*, due to the wide range of compounds of biotechnological interest that are produced from the intermediate p-coumaric acid; these include for example the flavonoid naringenin or the stilbene resveratrol[11,29]. For *B. subtilis* we optimized the production of a sialidase, as their enzymatic activity gained interest in relation to treatment of spinal cord injuries and there was a recent attempt to improve their activity and production using *B. subtilis*[30]. Both genes were cloned into their vectors using traditional restriction enzyme based cloning and were already expressed with high yield. After creating a combinatorial library that modifies the TIR, we could select variants that produced 2-8 fold more protein than the original clones. Most importantly, the sialidase production level in *B. subtilis* increased from 1 to 2 g/L in a 50 mL shake flask experiment, demonstrating that the specific TIR optimization and selection tool can achieve high yields, in the range of industrial titer levels, even when carried out in low density cultures. A substantial increase in yield is expected in industrial fed-batch fermentations.

The tool we developed has the potential to be easily translatable to other less amenable members of the *Bacillus* or *Lactococcus* families. With decreasing cost of DNA synthesis and the increasing interest in Gram-positive cell factories, different types of combinatorial libraries can be coupled with this selection system to screen for variants that produce a protein. Some examples are error prone PCR, DNA shuffling or any other in vitro and in vivo methodologies that generate genetic variability. The screening process is simple, inexpensive and can be carried out in a few days in any laboratory and therefore presents an attractive alternative to advanced screening methodologies.

Conclusion

We developed a selection-based system to screen for synthetically evolved TIRs Gram-positive hosts. The system responds linearly to increasing concentrations of antibiotic by preventing growth of non optimized library variants. Using our selection tool, we demonstrate improved expression of industrially relevant proteins in two cell factory hosts, namely TAL in *L. lactis* and a sialidase in *B. subtilis*.

Methods

Bacterial strains, media and growth conditions

Bacterial strains used in this study are listed in Supplementary Table S1.

Lactococcus lactis strain NZ9000 Δ hsd was used for all experimental procedures. Cells were grown at 30 °C in M17 broth (BD Difco, San Jose, CA, USA) supplemented with 1% glucose (GM17), without shaking. Electro competent cells were prepared as previously described[33]. Cultures were supplemented with 5 μg/mL erythromycin or 5 μg/mL chloramphenicol to select for plasmid presence, unless otherwise stated. Cultures were induced with 1.5 ng/mL nisin, unless otherwise stated.

Bacillus subtilis strain SCK6 (1A976 http://www.bgsc.org) was used for expression experiments. Cloning, library construction and propagation of plasmids were performed in *E. coli* NEB5α (New England Biolabs, Ipswich, MA, USA). B. subtilis was grown in lysogeny broth (LB) at 37 °C shaking (250 rpm). The antibiotics neomycin (5 μg/mL), kanamycin (50 μg/mL) and erythromycin (5 μg/mL) were supplemented when necessary. Transformation of integration vectors into B. subtilis was performed with chemically competent cells as previously described[34]. Correct genome integrations were confirmed by colony PCR and sequencing.

Plasmids and strain construction

Plasmids and oligonucleotides used in this study are listed in Supplementary Tables S2 and S3. All constructs were made with uracil excision cloning as previously described[35].

Plasmids used in *L. lactis* are based on pNZ8048 from the NIsin Controlled gene Expression (NICE) system[26] and expression was controlled by the inducible nisin promoter. Translationally coupled versions were constructed using the standard chloramphenicol acetyltransferase resistance gene derived from pNZ8048[26]. Hairpins were introduced by USER cloning using overlapping oligonucleotide tails coding for the hairpin[33,36]. The resulting plasmids pNZ-tal-hp-CmR was built by adding the hp-CmR module to the pNZ_FjTAL described previously[29]. The plasmid pNZ-*gfp*-hp-*CmR* was built in two steps: first by cloning a GFP folding reporter[37] into pNZ8048, using primers 1 and 2. In a second step the module hp-CmR was added to the vector.

B. subtilis plasmids were constructed in *E. coli* NEB5α (New England Biolabs, Ipswich, MA, USA), purified and subsequently transformed and integrated into *B. subtilis*. Integration vectors were based on the pDG268 plasmid, with integration in the *amyE* locus. Transcription was controlled by

four different constitutive promoters of increasing strength: P_{J23101}, P_{liaG}, P_{lepA}, and P_{veg}[27]. All four promoter variants were used to construct pDG-*gfp*-hp-*CmR*. Promoters were amplified from the *B. subtilis* genome. The synthetic promoter P_{J23101} was inserted into the plasmid using two overlapping oligonucleotides by PCR.

The replicating vector pDP66K-SIA-hp-CmR was constructed by adding the hp-CmR module to the plasmid pDP66K-Mv[30] which uses the promoter P_{32} to drive transcription.

Library construction

For the construction of TIR libraries a degenerated forward oligonucleotide specific for the gene of interest was designed. The six nucleotides upstream of the start codon were changed to all possible combinations whereas the six nucleotides downstream the start codon were changes to all possible synonymous codons.

Libraries for *L. lactis* were constructed by amplification of the whole pNZ-derived plasmid using the degenerated forward oligonucleotide and a reverse oligonucleotide with a pairing USER cloning overlap. The plasmid library was built by amplifying the template plasmid containing *gfp*, *gfp*-hp-*CmR* or *tal*-hp-*CmR* with the degenerate oligonucleotides and circularized using USER cloning as described elsewhere[35]. Libraries were transformed directly into *L. lactis* with no intermediate steps. The reference construct for the *tal* gene, here referred to as the TIR^{orig} clone, was previously described[29].

Libraries for *B. subtilis* were constructed in *E. coli* MC1061 by amplification of the whole pDG268neo or pDP66K-Mv plasmids using degenerated forward oligonucleotides and reverse oligonucleotides sharing 15 nucleotide homology with the forward oligonucleotide. Q5 polymerase (New England Biolabs, Ipswich, MA, USA) was used to amplify the template plasmid containing *gfp*, *gfp*-hp-*CmR* or *sia*-hp-*CmR*. Library construction was performed as described before[14]. The reference construct for the *sia* gene, here referred to as the TIR^{orig} clone, was previous described[30].

Expression and selection

Expression of individual *L. lactis* clones, were assayed using overnight cultures prepared by inoculating a single colony in 5 mL GM17 supplemented with respective antibiotics and incubated at 30 °C without shaking. Cultures were then back-diluted (1:50) into 5 mL of GM17 media

containing the appropriate antibiotics and incubated at 30 °C without shaking. At OD_{600} 0.3-0.6, cultures were induced with 1.5 ng/mL nisin and incubated for 3 h.

For the assessment of individual clones, *B. subtilis* overnight cultures were prepared by inoculating a single colony in 5 mL LB media supplemented with the appropriate antibiotics and incubated at 37 °C with shaking. Cultures were then back-diluted (1:50) into 5 mL of LB media containing the appropriate antibiotics and incubated at 37 °C with shaking for 5 h or 23 h.

Ca. 1 μg of plasmid library was transformed into *L. lactis* or *B. subtilis* using standard protocols[33,34]. Cells were recovered for 1h after transformation in GM17MC or LB media, transferred to GM17 or LB media supplemented with antibiotics and grown overnight at 30 °C for *L. lactis* and 37 °C for *B. subtilis*.

L. lactis cultures were then back-diluted (1:50) into 10 mL GM17 media containing the appropriate antibiotics and incubated at 30 °C. Cultures were induced at OD_{600} 0.3-0.6 and after 5h 0.2 OD units were plated on GM17 plates with increasing concentrations of chloramphenicol and incubated at 30 °C overnight.

After overnight incubation, B. subtilis cultures were back-diluted (1:50) into 5 mL LB media containing the appropriate antibiotics and incubated at 37 °C with shaking. 5 h after dilution, OD_{600} was measured and 0.2 OD units of cells were then plated on LB agar plates containing different concentrations of chloramphenicol and incubated overnight at 37 °C.

Selection was performed as previously described[36]. Selected expression variants were sequenced.

MIC determination

For *L. lactis* MIC determinations, 5 mL GM17 media with 5μg/mL erythromycin were inoculated with a single colony containing pNZGFP-hp-CmR and grown overnight at 30 °C. Cultures were then back-diluted (1:50) into 10 mL GM17 media supplemented with 5μg/mL erythromycin, and their growth monitored. At OD 0.3 the culture was split into 8 different 2 mL eppendorf tubes and induced with different concentrations of nisin (0; 0.25; 0.5; 0.75; 1; 1.5; 5 or 10 ng/mL). Cultures were incubated at 30 °C for 2 h and 0.01 ODU of each culture was transferred to a 96 wells plate (Greiner, Kremsmünster, Austria) containing 200 μL GM17 media (ca. 5x106 cfu/mL), and a serial dilution of chloramphenicol and respective concentrations of nisin as inducer. Plates were incubated for 15 h at 30 °C.

B. subtilis MIC determinations were performed in 5 mL LB media that were inoculated with 4 strains that constitutively expressed gfp-hp-CmR and grown overnight at 37 °C with shaking.

Cultures were then back-diluted (1:100) into 5 mL LB media containing the appropriate antibiotic in a 24-deep well plate (EnzyScreen, Heemstede, Netherlands). Cultures were incubated at 37 °C with shaking. After 2 h, 10 µL of each culture was transferred to a 96 well plate (Greiner, Austria) containing 100 µL LB media (ca. 5x105 cfu/mL) and a serial dilution of chloramphenicol. Plates were incubated for 18 h.

To assess translational coupling, fluorescence (Ex: 485 nm, Em: 516 nm for GFP) and OD_{600} were measured in an MX plate reader (Biotek, Winooski, VT, USA). The MIC value was defined as the lowest antibiotic concentration at which the final OD_{600} represented less than 10% of the entire population after background correction. Each MIC experiment was conducted with biological triplicates.

Protein detection and quantification

Western Blot analysis was performed by resuspending *L. lactis* grown as described above in 50% volume of CelLytic B (Sigma Aldrich, St. Louis, MO, USA) supplemented with lysozyme, egg white (Amresco, Solon, OH, USA), benzonase nuclease (≥250 units/µL, Sigma Aldrich, St. Louis, MO, USA) and Roche cOmplete™ Protease Inhibitor Cocktail (Sigma Aldrich, St. Louis, MO, USA) and incubated for 1 h before the samples were sonicated. An aliquot of 20µL of each sample was incubated for 1 hour with 1µL of a 1:10 dilution of CY5 dye (Amersham quick stain, GE healthcare, Chicago, IL, USA) for total protein quantification. The sample was then mixed with the same volume of 5x reducing sample buffer and heated to 95°C for 5 min for protein denaturation. 0.05 ODU of the samples were loaded onto a 4-20 % Mini-PROTEAN-TGX gel (BioRad, Hercules, CA, USA) and run for 35 min at 175 V.

For *B. subtilis* 10 µL of supernatant were mixed with 5 µL of 5x reducing sample buffer and heated to 95°C for 5 min for protein denaturation. 10 µL of the samples were loaded onto a 4-20 % Mini-PROTEAN-TGX gel (BioRad, Hercules, CA, USA) and run for 35 min at 175 V.

Proteins were transferred from the protein gel to a nitrocellulose membrane using the iBlot® dry blotting system (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) at 25 V for 7 min. The proteins were detected with the help of antigen-specific antibodies.

For *B. subtilis* sialidase an anti-His antibody (1:1000; Merck Millipore, Merck KGaA, Darmstadt, Germany) was used. The antibody was diluted in 5% w/v skim milk in TBS-T (20 mM Tris-HCl pH

7.6, 150 mM NaCl, 0.1 % v/v Tween-20), the secondary antibody was diluted in TBS-T. For *L. lactis* an anti-STREP (1:10000, Biorad, Hercules, CA, USA) directly coupled to HRP was used.

The HRP-coupled antibody was visualized using Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare, Chicago, IL, USA). The chemoluminescence signal was detected using a G:Box bioimager (Syngene, Cambridge, UK). The resulting images were analysed by densitometry using the Fiji software[38].

Sialidase protein purification

50 mL LB broth supplemented with 50 μg/mL kanamycin was inoculated with an overnight culture of *B. subtilis* SCK6 transformed with pDP66K-SIA-hp-CmR to an OD_{600} of 0.05. Cells were grown for 23 hours at 37°C with shaking (250 rpm). After 23 hours the supernatant was harvested by centrifugation at 5,000 x g for 15 min and 4°C and passed through a 0.45 μm filter (Frisenette ApS, Knebel, Denmark) and a 0.20 μm filter (Sartorius AG, Göttingen, Germany). The filtered supernatant was then concentrated using a 10K Amicon concentrator (Merck, Darmstadt, Germany), mixed with 10 mL purification buffer (50 mM NaH2PO4, 300 mM NaCl, 10 mM imidazole, pH 8.0) and again concentrated. The final concentrate was subjected to a Ni-NTA spin column (Qiagen, Hilden, Germany). The column was washed twice with 600 μL of washing buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0) and finally protein was eluted twice with 300 μL elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 500 mM imidazole, pH 8.0). To reduce imidazole concentration several cycles of concentrating and diluting in storage buffer (20 mM NaH2PO4, 100 mM NaCl and 10% glycerol, pH 7.4) were performed using a 10K Amicon concentrator (Merck, Darmstadt, Germany). The final volume was adjusted to 200 μL and protein concentration was estimated using BCA assay (Merck, Darmstadt, Germany). BSA was used as standard.

Sialidase activity assay

To determine activity of the sialidase, supernatant and concentrated supernatant of expression cultures of the original and optimized clones and purified proteins were diluted in 50 mM phosphate-citrate buffer pH 7.0. The enzymatic reaction was started by addition of the substrate pNP-Neu5Ac (Sigma Aldrich, St. Louis, MO, USA) at a concentration of 0.75 mM in a 100 μL reaction. Absorbance at 410 nm was monitored continuously for 1 h in an MX plate reader (Biotek, USA). The reaction rate was calculated from the slope of the initial linear section of the curve.

TAL activity assay

The TAL activity was measured as described by Jendersen and colleagues[29]. Briefly, expression of tal was induced with 1.5 ng/mL nisin in chemically defined media (CDM) as described above. Cultures were grown at 30°C for 16 hours, harvested at 13000g and the supernatant was recovered. The concentration of p-coumarate was measured by HPLC using a gradient method with two solvents (0.1% ammonium formate and acetonitrile) and quantified measuring absorbance at 290 nm.

List of abbreviations

Cm chloramphenicol

HRP horse radish peroxidase

LB lysogeny broth

NICE NIsin Controlled gene Expression

OD600 Optical density at 600 nm

SIA sialidase

SD Shine Dalgarno

TAL tyrosine ammonia lyase

TIR translation initiation region

Declarations

Funding

This work was supported by the Novo Nordisk Foundation and partially by the European Union's Seventh Framework Programme (BacHBerry, Project No. FP7-613793).

Authors' contributions

RF, MR, DOD and MHHN designed the experiments. RF performed the experiments in *L. lactis*. MR, RF and CHR performed the experiments in *B. subtilis*. RF and MR wrote the manuscript with contributions from all authors. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to acknowledge Carsten Jers for discussion and advice in the *B. subtilis* part of this study and sharing of plasmids or strains when needed, and Alexey Dudnik for general discussion on *L. lactis* troubleshooting and sharing of plasmids or strains when needed.

Competing interests

The authors declare possible competing interests. RF, MR, DOD and MHHN have submitted a patent application on the use of the hairpin coupling mechanism in prokaryotes. DOD has submitted a patent on the TIR optimization. CHR declares no competing financial interests.

Availability of data and materials

All material available upon request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Cherry JR, Fidantsef AL. Directed evolution of industrial enzymes: An update. Curr. Opin. Biotechnol. 2003;14:438–43.
2. Gavrilescu M, Chisti Y. Biotechnology - A sustainable alternative for chemical industry. Biotechnol. Adv. 2005;23:471–99.
3. Ferrer-Miralles N, Villaverde A. Bacterial cell factories for recombinant protein production; expanding the catalogue. Microb. Cell Fact. [Internet]. 2013;12:113. Available from: http://microbialcellfactories.biomedcentral.com/articles/10.1186/1475-2859-12-113
4. Chen R. Bacterial expression systems for recombinant protein production: E. coli and beyond. Biotechnol. Adv. [Internet]. Elsevier Inc.; 2012;30:1102–7. Available from: http://dx.doi.org/10.1016/j.biotechadv.2011.09.013
5. van Dijl JM, Hecker M. Bacillus subtilis: from soil bacterium to super-secreting cell factory. Microb. Cell Fact. [Internet]. 2013;12:3. Available from: http://microbialcellfactories.biomedcentral.com/articles/10.1186/1475-2859-12-3
6. Song AA-L, In LLA, Lim SHE, Rahim RA. A review on Lactococcus lactis: from food to factory. Microb. Cell Fact. [Internet]. BioMed Central; 2017;16:55. Available from: http://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-017-0669-x
7. Kunji ERS, Slotboom DJ. Lactococcus lactis as host for overproduction of functional membrane proteins. Biochim. Biophys. Acta (BBA … [Internet]. 2003; Available from: http://www.sciencedirect.com/science/article/pii/S0005273602007125

8. Chan KAWAI, Slotboom D. Functional expression of eukaryotic membrane proteins in Lactococcus lactis. 2005;3048–56.
9. Morello E, Bermúdez-Humarán LG, Llull D, Solé V, Miraglio N, Langella P, et al. Lactococcus lactis, an efficient cell factory for recombinant protein production and secretion. J. Mol. Microbiol. Biotechnol. [Internet]. 2008 [cited 2014 Sep 26];14:48–58. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17957110
10. Le Loir Y, Azevedo V, Oliveira SC, Freitas DA, Miyoshi A, Bermudez-Humaran LG, et al. Protein secretion in Lactococcus lactis : an efficient way to increase the overall heterologous protein production. Microb Cell Fact [Internet]. 2005;13:1–13. Available from: https://doi.org/10.1186/1475-2859-4-2
11. Dudnik A, Almeida AF, Andrade R, Avila B, Bañados P, Barbay D, et al. BacHBerry: BACterial Hosts for production of Bioactive phenolics from bERRY fruits. Phytochem. Rev. [Internet]. 2017; Available from: http://link.springer.com/10.1007/s11101-017-9532-2
12. Song AA, Abdullah JO, Abdullah MP, Shafee N, Rahim RA. Functional expression of an orchid fragrance gene in Lactococcus lactis. Int J Mol Sci [Internet]. 2012;13. Available from: https://doi.org/10.3390/ijms13021582
13. Martinez-Cuesta MC, Gasson MJ, Narbad A. Heterologous expression of the plant coumarate: coA ligase in Lactococcus lactis. Lett Appl Microbiol [Internet]. 2005;40. Available from: https://doi.org/10.1111/j.1472-765X.2004.01621.x
14. Mirzadeh K, Martínez V, Toddo S, Guntur S, Herrgård MJ, Elofsson A, et al. Enhanced Protein Production in Escherichia coli by Optimization of Cloning Scars at the Vector-Coding Sequence Junction. ACS Synth. Biol. 2015;4:959–65.
15. Laursen BS, Sørensen HP, Mortensen KK, Sperling-Petersen HU. Initiation of protein synthesis in bacteria. Microbiol. Mol. Biol. Rev. [Internet]. 2005;69:101–23. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-14844340954&partnerID=tZOtx3y1
16. Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LBA, Gill RT. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. Nat. Biotechnol. [Internet]. Nature Publishing Group; 2010;28:856–62. Available from: http://www.nature.com/doifinder/10.1038/nbt.1653
17. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, et al. Programming cells by multiplex genome engineering and accelerated evolution. Nature [Internet]. Nature Publishing Group; 2009;460:894–8. Available from: http://dx.doi.org/10.1038/nature08187
18. Sjostrom SL, Bai Y, Huang M, Liu Z, Nielsen J, Joensson HN, et al. High-throughput screening for industrial enzyme production hosts by droplet microfluidics. Lab Chip [Internet]. 2014;14:806–13. Available from: http://xlink.rsc.org/?DOI=C3LC51202A
19. Brendan P. Cormack; Raphael H. Valdivia; Stanley Falkow. FACS-optimized mutants of the green fluorescent protein (GFP). Gene. 1996;173:33–8.
20. Yang G, Withers SG. Ultrahigh-throughput FACS-based screening for directed enzyme evolution. ChemBioChem. 2009;10:2704–15.
21. Dietrich JA, McKee AE, Keasling JD. High-Throughput Metabolic Engineering: Advances in Small-Molecule Screening and Selection [Internet]. Annu. Rev. Biochem. 2010. Available from: http://www.annualreviews.org/doi/10.1146/annurev-biochem-062608-095938
22. Chen X, Zaro JL, Shen WC. Fusion protein linkers: Property, design and functionality. Adv. Drug Deliv. Rev. [Internet]. Elsevier B.V.; 2013;65:1357–69. Available from: http://dx.doi.org/10.1016/j.addr.2012.09.039
23. Linares DM, Geertsma ER, Poolman B. Evolved Lactococcus lactis strains for enhanced expression of recombinant membrane proteins. J. Mol. Biol. [Internet]. Elsevier Ltd; 2010 [cited 2014 Nov 11];401:45–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20542040
24. Schallmey M, Frunzke J, Eggeling L, Marienhagen J. Looking for the pick of the bunch: High-throughput screening of producing microorganisms with biosensors. Curr. Opin. Biotechnol. [Internet]. Elsevier Ltd; 2014;26:148–54. Available from: http://dx.doi.org/10.1016/j.copbio.2014.01.005
25. Rennig M, Martinez V, Mirzadeh K, Dunas F, Röjsäter B, Daley DO, et al. TARSyn: Tuneable antibiotic resistance devices enabling bacterial synthetic evolution and protein production. In Preparation. 2017;

26. Kuipers OP, De Ruyter PGGA, Kleerebezem M, De Vos WM, Ruyter P, Kleerebezem M, et al. Quorum sensing-controlled gene expression in lactic acid bacteria. J Biotechnol [Internet]. 1998;64:15–21. Available from: https://doi.org/10.1016/S0168-1656(98)00100-X
27. Radeck J, Kraft K, Bartels J, Cikovic T, Dürr F, Emenegger J, et al. The Bacillus BioBrick Box: generation and evaluation of essential genetic building blocks for standardized work with Bacillus subtilis. J. Biol. Eng. [Internet]. 2013;7:29. Available from: http://jbioleng.biomedcentral.com/articles/10.1186/1754-1611-7-29
28. Jers C, Kobir A, Søndergaard EO, Jensen PR, Mijakovic I. Bacillus subtilis Two-Component System Sensory Kinase DegS Is Regulated by Serine Phosphorylation in Its Input Domain. Kao KC, editor. PLoS One [Internet]. 2011;6:e14653. Available from: http://dx.plos.org/10.1371/journal.pone.0014653
29. Jendresen CB, Stahlhut SG, Li M, Gaspar P, Siedler S, Förster J, et al. Highly active and specific tyrosine ammonia-lyases from diverse origins enable enhanced production of aromatic compounds in bacteria and Saccharomyces cerevisiae. Appl. Environ. Microbiol. 2015;81:4458–76.
30. Jers C, Guo Y, Kepp KP, Mikkelsen JD. Mutants of micromonospora viridifaciens sialidase have highly variable activities on natural and non-natural substrates. Protein Eng. Des. Sel. 2015;28:37–44.
31. Singh R, Kumar M, Mittal A, Mehta PK. Microbial enzymes: industrial progress in 21st century. 3 Biotech. Springer Berlin Heidelberg; 2016;6:1–15.
32. Mendez-Perez D, Gunasekaran S, Orler VJ, Pfleger BF. A translation-coupling DNA cassette for monitoring protein translation in Escherichia coli. Metab. Eng. [Internet]. Elsevier; 2012;14:298–305. Available from: http://dx.doi.org/10.1016/j.ymben.2012.04.005
33. Holo H, Nes IF. High-Frequency Transformation, by Electroporation, of Lactococcus lactis subsp. cremoris Grown with Glycine in Osmotically Stabilized Media. Appl. Environ. Microbiol. [Internet]. Am Soc Microbiol; 1989;55:3119. Available from: http://aem.asm.org.globalproxy.cvt.dk/cgi/content/abstract/55/12/3119
34. Zhang XZ, Zhang YHP. Simple, fast and high-efficiency transformation system for directed evolution of cellulase in Bacillus subtilis. Microb. Biotechnol. [Internet]. 2011;4:98–105. Available from: http://doi.wiley.com/10.1111/j.1751-7915.2010.00230.x
35. Cavaleiro AM, Kim SH, Seppälä S, Nielsen MT, Nørholm MHH. Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning. ACS Synth. Biol. 2015;4:1042–6.
36. Rennig M, Daley DO, Nørholm MHH. Selection of Highly Expressed Gene Variants in Escherichia coli Using Translationally Coupled Antibiotic Selection Markers [Internet]. 1st ed. Jensen MK, Keasling JD, editors. Synth. Metab. Pathways Methods Protoc. New York, NY: Humana Press part of Springer Science & Business Media; 2018. Available from: http://www.springer.com/gp/book/9781493972944
37. Waldo GS, Standish BM, Berendzen J, Terwilliger TC. Rapid folding assay using green fluorescent protein. 1999;691–5.
38. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: An open source platform for biological image analysis. Nat. Methods. 2012;9:676–82.

Supplementary material for

A synbio approach for selection of highly expressed gene variants in Gram-positive bacteria

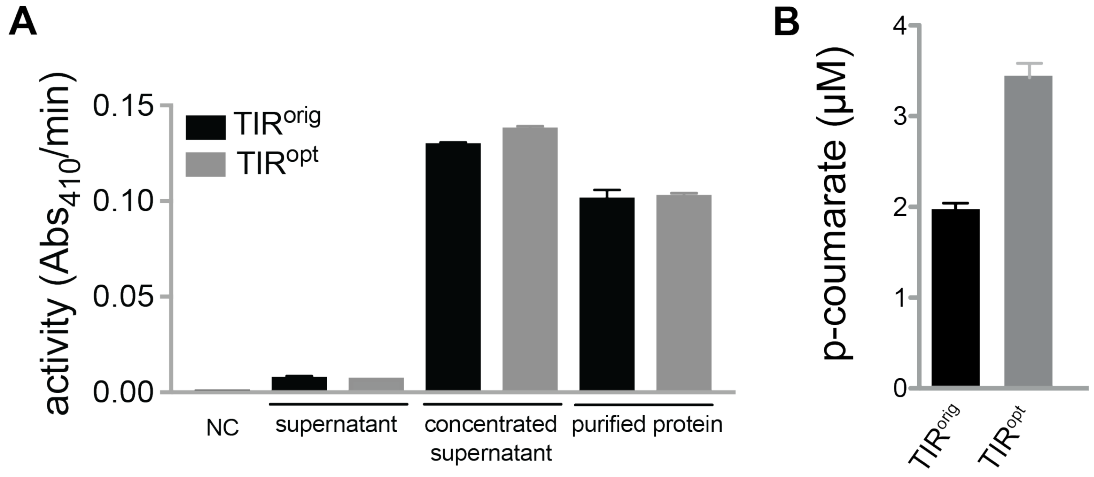


Figure S1: Activity assessment of selected library clones. To assure that optimized variants are still active, activity assays for sialidase and TAL were performed. (A) Activity of the optimized sialidase clone was measured as absorbance at 410 nm per minute. Activity for culture supernatant, concentrated culture supernatant and purified protein were assessed. pNP-Neu5Ac was used as substrate. The negative control (NC) contained buffer and substrate but no enzyme was supplemented. (B) Activity of the optimized TAL clone was measured as product formation in µM by HPLC.

Table S1. Strains used in this study

Strain	Genotype	Source/Reference
L. lactis NZ9000 - Δ *hsd*	Δ *hsd, pepN::nisRK*	Dudnik, not published
B. subtilis SCK6	Em, his, nprE18, aprE3, eglSΔ102, bglT/bglSΔEV, lacA::PxylA-comK	1
B. subtilis SCK6 Pveg-*gfp*	Em, his, nprE18, aprE3, eglSΔ102, bglT/bglSΔEV, lacA::PxylA-comK, amyE::Pveg-gfp-hp-CmR	This study
B. subtilis SCK6 PlepA-*gfp*	Em, his, nprE18, aprE3, eglSΔ102, bglT/bglSΔEV, lacA::PxylA-comK, amyE::PlepA-gfp-hp-CmR	This study
B. subtilis SCK6 PliaG-*gfp*	Em, his, nprE18, aprE3, eglSΔ102, bglT/bglSΔEV, lacA::PxylA-comK, amyE::PliaG-gfp-hp-CmR	This study
B. subtilis SCK6 PJ23101-*gfp*	Em, his, nprE18, aprE3, eglSΔ102, bglT/bglSΔEV, lacA::PxylA-comK, amyE::PJ23101-gfp-hp-CmR	This study
E. coli NEB5α	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	New England Biolabs, USA
E. coli MC1061	araD139, Δ(ara, leu)7697, ΔlacX74, galU-, galK-, hsr-, hsm+, strA	In house

Table S2. Plasmids used in this study

Plasmid	Property	Source/Reference
pNZ8048	*L. lactis* gene expression vector, Cm^R, PnisA	2
pNZ_FjTAL	Pnz8048 vector with PnisA-*tal* insert - Cm^R	3
pNZ-TAL-hp-Cm^R	Pnz8048 vector Erm^R, with PnisA-*tal*-hp-Cm^R insert	This study
pNZ-GFP-hp-Cm^R	Pnz8048 vector, Erm^R, with PnisA-*gfp*-hp-Cm^Rinsert	This study
pDG268-neo	Shuttle vector, ColE1 origin, Amp^R, MCS for integration into *amyE* locus with Neo^R	4
pDG-Pveg-GFP-hp-Cm^R	Shuttle vector, ColE1 origin, Amp^R, for integration of P_{veg}-gfp-hp-Cm^R into *amyE* locus	This study
pDG-PlepA-GFP-hp-Cm^R	Shuttle vector, ColE1 origin, Amp^R, for integration of P_{lepA}-gfp-hp-Cm^R into *amyE* locus	This study
pDG-PliaG-GFP-hp-Cm^R	Shuttle vector, ColE1 origin, Amp^R, for integration of P_{liaG}-gfp-hp-Cm^R into *amyE* locus	This study
pDG-PJ23101-GFP-hp-Cm^R	Shuttle vector, ColE1 origin, Amp^R, for integration of P_{J23101}-gfp-hp-Cm^R into *amyE* locus	This study
pDP66K-Mv	Cloning and expression vector, Km^R, encodes P32-sp-*sia*	5
pDP66K-SIA-hp-Cm^R	Cloning and expression vector, Km^R, encodes P32-sp-*sia*-hp-Cm^R	This study

Table 3. Oligonucleotides used in this study

Name	Sequence (5' --> 3')
Cloning oligonucleotides	
2995_pDGneo_fwd	AAAAGCAUTAGTGTATCAACAAGC
2989_pDGneo_rev	AGCTATTUCAGCTGCGCTTTTTCCATTATGTACTATTTCGATCAGAC
2993_CmR_fwd	ATAGGAGGUCCTCCTatgTCAaactttaataaaattgatttagacaattg
2994_CmR_rev	AAATAGCUGCGCTTTTTTGTGTCATAAttataaaagccagtcattaggc
2999_pVeg_fwd	ATGCTTTUGGAGTTCTGAGAATTGGTATG
3000_pVeg_rev	ACAGTAGUACTACATTTATTGTACAACACGAGC
3001_pLepA_fwd	ACAGTAGUACTATTAAACGCAAAATACACTAGC
3002_pLepA_rev	ATGCTTTUAGTCAATGTATGAATGGATACGG
3003_pLiaI_fwd	ACAGTAGUTCGTTTTCCTTGTCTTCATCT
3004_pLiaI_rev	ATGCTTTUATTGGCCAAAGCAGAAAG
3005_pLiaG_rev	ATGCTTTUCAAAAATCAGACCAGACAAAAG
3006_pLiaG_fwd	ACAGTAGUTCATTCATTCTATTATAAAGGAAAAGC
3044_pJ23101_fwd	ACAGTAGUgctagcataatacctaggactgagctagctgtaaaGGATCCTAGAAGCT TATCGA
3135_ GFPsf_rev	ACCTCCTAUGTCAttTgtatagttcatccatgcc
3136 _GFPsf_fwd	AGGatgcgUaaaggagaagaactt
3137_pDG_CmR_rev	ATAGGAGGUCCTCctatgtcaa
3138 _pDG_Pveg_fwd	acgcatCCUCGAGcctcctA
3206_pDG_pVeg-GFP-lib_fwd	ACTACTGUaggaggCNNNNNNatgcgNaaRggagaagaacttttcactgg
3474_pDP66K_fwd	AGGATCCUGCCTGCGAT

3475_sia-His_rev	ACCTCCTAUGTCAATGATGGTGGTGATGG
3476_Term_CmR_rev	AGGATCCUAAAAAGCGCAGCTGAAA
3489_pDP66K-P32-Sia-lib_fwd	ACGATTACATAGGAGGNNNNNNATGAARAARTTTCTGAAATCGACAGC TGC
3490_pDP66K-lib_rev	CTCCTATGTAATCGTTTGAATTCCAGGCTTGTCCGCTGTCGCCGGATC C
1_GFPfr_fwd	actcaccaUGTCCAAAGGAGAAGAACTT
2_GFPfr_rev	aacCTTTGUAGAGCTCATCCATGC
3a_hpCmR_rev	ACCTCCTAUGTCATTTTTCAAATTGTGGATGGC
3_CmR_fwd	ATAGGAGGUCCTCCTatgTCAaactttaataaaattgatttagacaattg
4_CmR+histag_rev	ATGATGAUGGGCCGCAAGCTTtaaaagccagtcattaggcctatc
4b_Backbone_HIS_fwd	ATCATCAUCACCACCACCACCACtaatcaattgaaatggcaattaaac
5_RND-GFPfr_fwd	ataaattaUaaggaggcactcaccATGTCNAARGGAGAAGAACTTTTCACTGG
6_RND-GFP-fr_rev	ATAATTTAUTTTGTAGTTCCTTCG
7_ TAL_Strep_fwd	ATCCACAAUTTGAAAAATAAtctagagagctcaagctttct
8_ TAL_Strep_rev	ATTGTGGAUGGCTCCAGCTTTTAGAACCattgttaatcaggtggtcttttac
9_ ERMSwap_fwd	attgaaUGCTTCAGTTGTCTTATTTCTAGATCT
10_ERMSwap_rev	atctcataUTATTTCCTCCCGTTAAATAATAGATA
11_ BB-ERM_rev	attcaaUaatccctcctctca
12_ BB-ERM_fwd	atatgagaUaatgccgactgt
13_TALRndlib_fwd	ATAAATTAUAAGGAGGCANNNNNNatgAAYACNatcaacgaatatctgagcc
14_TALRndlib_rev	ATAATTTAUTTTGTAGTTCCTTCG

Sequencing oligonucleotides	
3052_AmyE_SeqS_rev	TGCCTGAACGAGAAGCTAT
3051_AmyE_SeqL_rev	TATATAAACCATTTAGCACGTAATCA
2990_pDGseq_rev	TGTATCAAGATAAGAAAGAACAAGTTC
3064_AmyE_SeqL_fwd	ATGTTTGCAAAACGATTCA
3015_pDGseq_fwd	CCAATGAGGTTAAGAGTATTCC
3528_P32_seq_fwd	gatatgataagattaatagt
3529_Sialidase_seq_rev	GTTGGGCGGCCGTCGTATGA
3530_KanR_seq_fwd	aagcctgattgggagaaaat
TalSeq_fwd	gattaccattgttcaggcg
TalSeq_rev	acaaaccggactcagcg
FW_insert_seq	actaacctgccccgttagt
RV insert seq	ATTCCTTGGUCCTTTAATTGGTGGACA

References

1. Zhang X, Zhang YP. Simple , fast and high-efficiency transformation system for directed evolution of cellulase in Bacillus subtilis. 2011;4:98–105.

2. Kuipers OP, De Ruyter PGGA, Kleerebezem M, De Vos WM, Ruyter P, Kleerebezem M, et al. Quorum sensing-controlled gene expression in lactic acid bacteria. J Biotechnol [Internet]. 1998;64:15–21. Available from: https://doi.org/10.1016/S0168-1656(98)00100-X

3. Jendresen CB, Stahlhut SG, Li M, Gaspar P, Siedler S, Förster J, et al. Highly active and specific tyrosine ammonia-lyases from diverse origins enable enhanced production of aromatic compounds in bacteria and Saccharomyces cerevisiae. Appl. Environ. Microbiol. 2015;81:4458–76.

4. Jers C, Kobir A, Søndergaard EO, Jensen PR, Mijakovic I. Bacillus subtilis Two-Component System Sensory Kinase DegS Is Regulated by Serine Phosphorylation in Its Input Domain. 2011;6.

5. Jers C, Guo Y, Kepp KP, Mikkelsen JD. Mutants of micromonospora viridifaciens sialidase have highly variable activities on natural and non-natural substrates. Protein Eng. Des. Sel. 2015;28:37–44.

Paper 4

SEVA Linkers: A Versatile and Automatable DNA Backbone Exchange Standard for Synthetic Biology

Kim S.H., Cavaleiro A.M., Rennig M., Nørholm, M.H.H.

Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hørsholm DK-2970, Denmark

ACS Synthetic Biology 2016, 5, 1177−1181

DOI: 10.1021/acssynbio.5b00257

Abstract

DNA vectors serve to maintain and select recombinant DNA in cell factories, and as design complexity increases, there is a greater need for well-characterized parts and methods for their assembly. Standards in synthetic biology are top priority, but standardizing molecular cloning contrasts flexibility, and different researchers prefer and master different molecular technologies. Here, we describe a new, highly versatile and automatable standard "SEVA linkers" for vector exchange. SEVA linkers enable backbone swapping with 20 combinations of classical enzymatic restriction/ligation, Gibson isothermal assembly, uracil excision cloning, and a nicking enzyme-based methodology we term SEVA cloning. SEVA cloning is a simplistic one-tube protocol for backbone swapping directly from plasmid stock solutions. We demonstrate the different performance of 30 plasmid backbones for small molecule and protein production and obtain more than 10-fold improvement from a four-gene biosynthetic pathway and 430-fold improvement with a difficult-to-express membrane protein. The standardized linkers and protocols add to the Standard European Vectors Architecture (SEVA) resource and are freely available to the synthetic biology community.

Keywords: *synthetic biology standards, plasmid backbone exchange, standard parts characterization, cell factory design*

Introduction

The early steps in engineering of microbial cell factories typically involve a choice of vector for gene expression. This vector enables maintenance (replication) of the genetic elements of interest, e.g., by containing an origin of replication or elements ensuring transfer to the microbial genome, and often includes a selectable trait, in bacteria typically in the form of antibiotic resistance. These initial choices can have a major impact on the performance of the cell factory, and balancing these factors is imperative to optimize production.[1,2]

Plasmids are extrachromosomal DNA elements that are nonessential, can replicate autonomously and are easily modified in vitro and thus represent an extremely powerful toolbox for molecular biology. The number of plasmid molecules in a single cell specifies the amount of gene copies available for expression, and this copy number is determined by different genetic elements at the origin of replication.[3] Toxicity of plasmid-encoded proteins is usually the highest metabolic burden

for a production host,[4] and gene overexpression is a stress for the organism that needs to cope with this metabolic overload and prevent the system's breakdown.[2] Moreover, extra DNA elements and gene expression will likely compete with the native DNA for essential resources, e.g., the native RNA polymerase.[5,6]

In a typical laboratory setup, horizontal gene transfer events are selected for with the aid of antibiotic resistance genes. Antibiotics either inhibit bacterial cell growth (bacteriostatics) or cause bacterial cell death (bactericidals).[7] Thus, antibiotic selection is inherently linked to metabolic burden, fitness costs and physiological changes [8–11]

Most synthetic biology and metabolic engineering projects require a first step of DNA assembly,[12–14] and with increasingly advanced design requirements, simple methods for genetic elements exchange are highly attractive. Struggling to find the best-suited assembly and exchange strategies is common among researchers since all of the described DNA cloning methods possess different limitations. Moreover, a consensus method- ology is hard to agree on: best exemplified with the paradoxical high number of assembly strategies available at the Registry of Standard Biological Parts.[15] Nevertheless, standardized genetic parts and methods for their assembly are very important for continuous progress in the synthetic biology field.[16,17] For example, standards surely will enable more systematic and reliable approaches to assay the performance and robustness of genetic elements.

Here we attempt to address the standardization paradox by designing small multifunctional DNA fragments designed to link together genetic elements often used in bacterial cell factories. The resulting linkers contain elements enabling continuous reassembly with several of the most common DNA assembly methodologies (e.g., restriction enzymes, Gibson assembly and uracil excision cloning) and with a new extraordinary simple protocol for plasmid backbone exchange. Finally, we demonstrate the usefulness of this resource by systematically comparing bacterial production of the membrane protein NarK and the food coloring pigment β-carotene produced from a four-gene biosynthetic pathway, each with a total of 30 combinations of origins of replication and antibiotic resistance markers.

Results and Discussion

In the Standard European Vector Architecture (SEVA),[18,19] different rare restriction sites flank three basic genetic elements: (1) antibiotic selection markers, (2) origins of replication and (3) the so-called cargo that contains the genetic elements necessary for the end-application of the cell factory (very often a promoter driving expression of a gene, Figure S1). Importantly, these genetic elements have well-defined borders that are useful for parts exchange and a catalogue of all 54 combinations of nine origins of replication with six antibiotic resistance markers is available.

To make this great resource compatible with a range of state- of-the-art DNA assembly methods, we designed two multi- functional SEVA linker sequences that hosted five and four additional rare restriction sites, respectively (PacI, NotI, AscI, SgrDI, MauBI on one side and FseI, Sbf II, MreI, SpeI on the other side, Figure 1a). This enables backbone swapping with a total of 20 combinations of these enzymes. Additionally, all cargo elements, flanked with these linkers, are easily inserted in the existing SEVA system using the outermost PacI and SpeI sites (but violates the SEVA design rules by reusing the FseI and AscI sites, Figure S1). The saturation with rare restriction sites means that it should always be possible to find a suitable pair of restriction enzymes for backbone swapping. In other words this minimizes the consequence of "forbidden sites" in the cargo. Moreover, these unique linkers sequences make it possible to design highly specific oligonucleotides (for examples see Supporting Information Table S1) that anneal in these regions, while hosting features compatible with state-of-the-art DNA assembly methods such as Gibson isothermal assembly[20] and uracil excision[21] (Supporting Information Table S1, Figure S2). Such oligonucleotides can be performance benchmarked and kept in the freezer for continuous reuse when assembling new parts with these technologies.

Another recently popularized assembly method is Golden Gate cloning.[22] One of the biggest advantages of Golden Gate cloning is that DNA can be exchanged directly from (compatible) plasmid stock solutions with a simple protocol, whereas a drawback is the frequent occurrence of the type IIS restriction sites typically used: "forbidden sites". Inspired by some of the features in Golden Gate cloning, we designed an extension to the SEVA linkers enabling backbone exchange with a very simple protocol directly from plasmid stocks. Instead of type IIS restriction sites, our design uses two pairs of nicking restriction sites (Figure 1a) that together form two different 7 bp cohesive ends (Figure 1a and Supporting Information Figure S3). We initially used the Nb.BtsI nicking enzyme, but changed to Nt.BbvCI because the recognition site is 7 bp and only occurs

rarely in standard sized DNA constructs. Furthermore, because the mutant enzyme only cuts one strand, occurrences outside the SEVA linker will probably not affect cloning efficiency significantly. This solves the forbidden site issue.

We designed two SEVA-linker-flanked cargo elements, one expressing lacZ and one expressing gfp, with different antibiotic selection markers, and checked their ability to recombine by mixing the plasmids together with the nicking enzyme and plating on media with the different antibiotics (Supporting Information Figure S4). Ultimately, the idea is that a cargo "donor" plasmid should be transferred to a backbone "acceptor" plasmid at high efficiency and specificity. To this end, we incorporated the toxic ccdB gene[23] in the backbone acceptor and selected recombinants by transforming constructs into standard (i.e., ccdB-incompatible) cloning strains with the backbone-defined antibiotic selection (Figure 1a). As an initial proof of concept, we examined gfp flanked by the two SEVA linkers in the commercially available pCDF backbone (CloDF13 origin and spectinomycin resistance) transferred to an acceptor plasmid containing ccdB in the pACYC backbone (p15A origin and chloramphenicol resistance). We then observed fluorescent colonies forming on chloramphenicol- containing plates with high efficiency (Figure 1b). Optimization of the ratio of donor and acceptor plasmids in the reaction increased efficiency approximately 2-fold, whereas adding T4 DNA ligase to the mixture enhanced efficiency more than 10- fold (Figure 1b). The optimized and simple protocol is described in detail in the Supporting Information. Notably, rather than getting it right the first time, several similar nicking enzyme-based designs were tested for efficiency and specificity before we settled on the sequence presented here (see Supporting Information Figure S5).

We based our further work on the comprehensive pSEVA collection as a backbone acceptor series. By combining five origins of replication (pBBR1: #3, p15A: #6, pSC101: #7, pUC: #8, pBR322/ROP: #9) and six antibiotic resistance markers (ampicillin: #1, kanamycin: #2, chloramphenicol: #3, spectino- mycin: #4, tetracycline: #5, gentamycin: #6), we created 30 different backbone acceptors with the counter-selection marker ccdB flanked by SEVA linkers as the initial cargo.

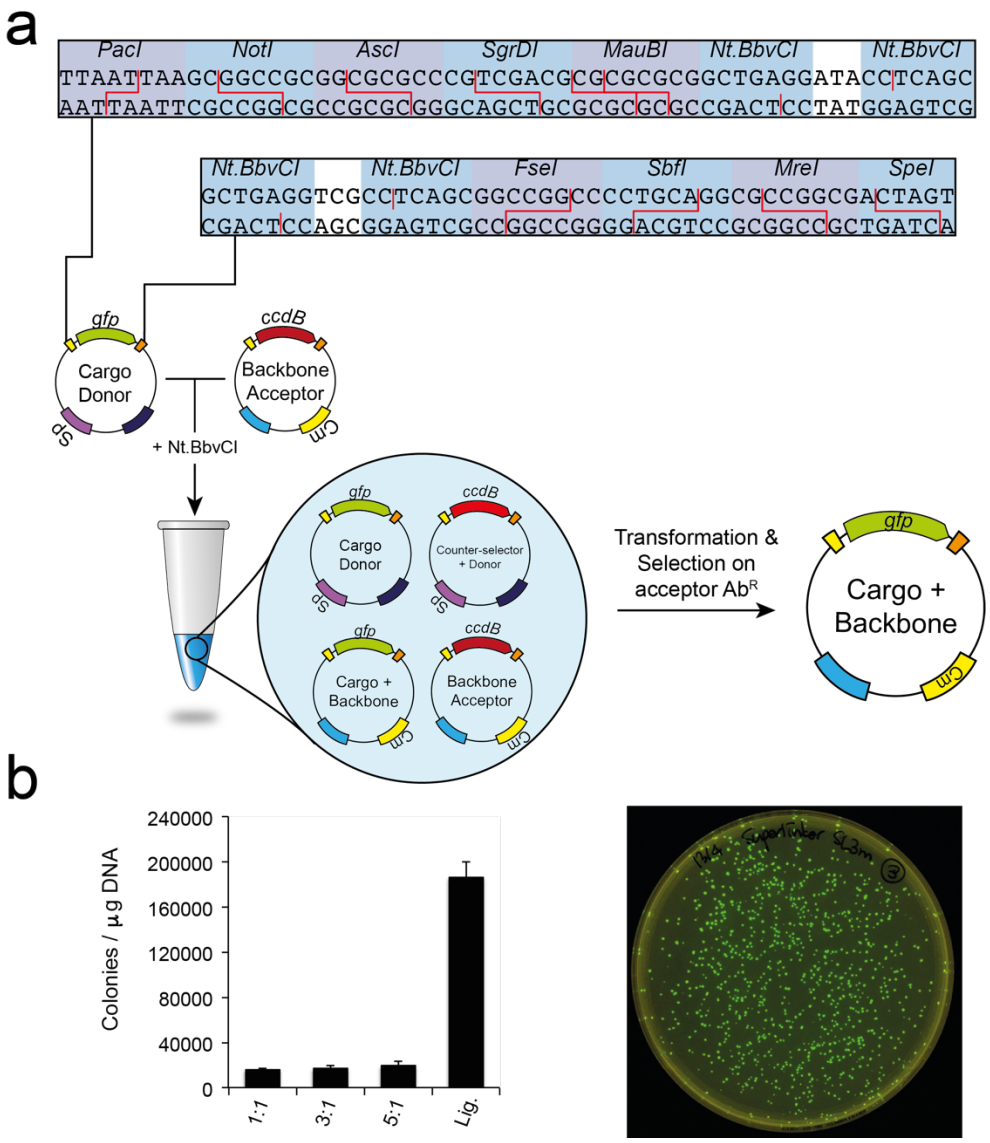


Figure 1. Illustration and optimization of SEVA-linker-based backbone exchange. (a) The cargo, here illustrated with a plasmid-encoded gfp, is flanked by two multifunctional linker sequences altogether hosting a total of nine rare restriction sites and four Nt.BbvcI nicking enzyme recognition sites (see expanded view). When mixed with the Nt.BbvcI enzyme and a backbone acceptor plasmid, hosting the toxic ccdB gene flanked with the same linkers, the recombined cargo can be selected on the antibiotic defined by the acceptor backbone. The expanded view of the reaction shows all assembly possibilities. (b) Left panel: efficiency of recombination with different ratios of backbone acceptor and cargo donor plasmids, and with the addition of T4 DNA ligase to the mixture. Right panel: representative picture showing the efficient recombination of a gfp-expressing cargo into a new backbone acceptor plasmid.

To assay the performance of our 30 standardized plasmid backbones in two typical, cell factory-type experimental settings, we swapped-in the four-gene *crtEBIY* biosynthetic pathway for β-carotene from *Pantoea ananatis* and the membrane protein-encoding *narK-gfp* in all 30 constructs. *crtEBIY* is industrially relevant and a convenient model pathway mainly due to the simple product output (orange color), but also because the robustness of the phenotype

seems sensitive to the cell factory design parameters.[24] Indeed, we were unable to obtain surviving colonies by swapping the pathway into vectors with the high-copy pUC origin of replication and we obtained highly variable phenotypes when T7 polymerase was used to drive expression from the construct in comparison with the weaker K1F variant (Figure 2a).[24–26] Overall, we observed a 10-fold difference in the β-carotene product titers, from the lowest to the highest performing cell factories (Supporting Information Table S3) and several of the combinations showed clear toxicity and population bias effects (e.g., pBR322/ROP in combination with chloramphenicol or pSC101 in combination with tetracycline, Figure 2a). In contrast, the pBR322/ROP origin in combination with both ampicillin and tetracycline was highly performing in both expression strains. With the membrane protein NarK,[27] the variation in expression yield was even more prominent (Figure 2b); the difference between the highest and the lowest performing combination of parts was in this case an impressive 430-fold (Supporting Information Table S3). Generally, using tetracycline selection, gentamycin selection or the high copy pUC origin had a negative impact on narK expression, whereas the p15A origin (low copy number) and spectinomycin selection seemed to positively impact the expression level.

Comparing the two different test cases, small molecule and protein production, showcases the value of a synthetic biology approach (i.e., systematic studies with standardized parts) and provides future design guidelines and a toolbox for similar experiments. For example, the high copy pUC origin of replication is likely a poor choice for anything but DNA production. The negative impact of the tetracycline selection may in contrast only reflect the fact that the resistance gene encodes a membrane protein that could compete for factors involved in membrane translocation important for production of NarK, and thus may be a particularly poor choice for membrane protein production. In many cases we observed clear population bias effects by simple visual inspection on agar plates, and the different robustness of the T7- and the K1F- based bacterial hosts highlights the value of adding promoter tuning as an extra dimension in the cell factory performance screen. These observations could be supported by an array of omics studies leading to a highly informed theoretical framework for rational cell factory design.

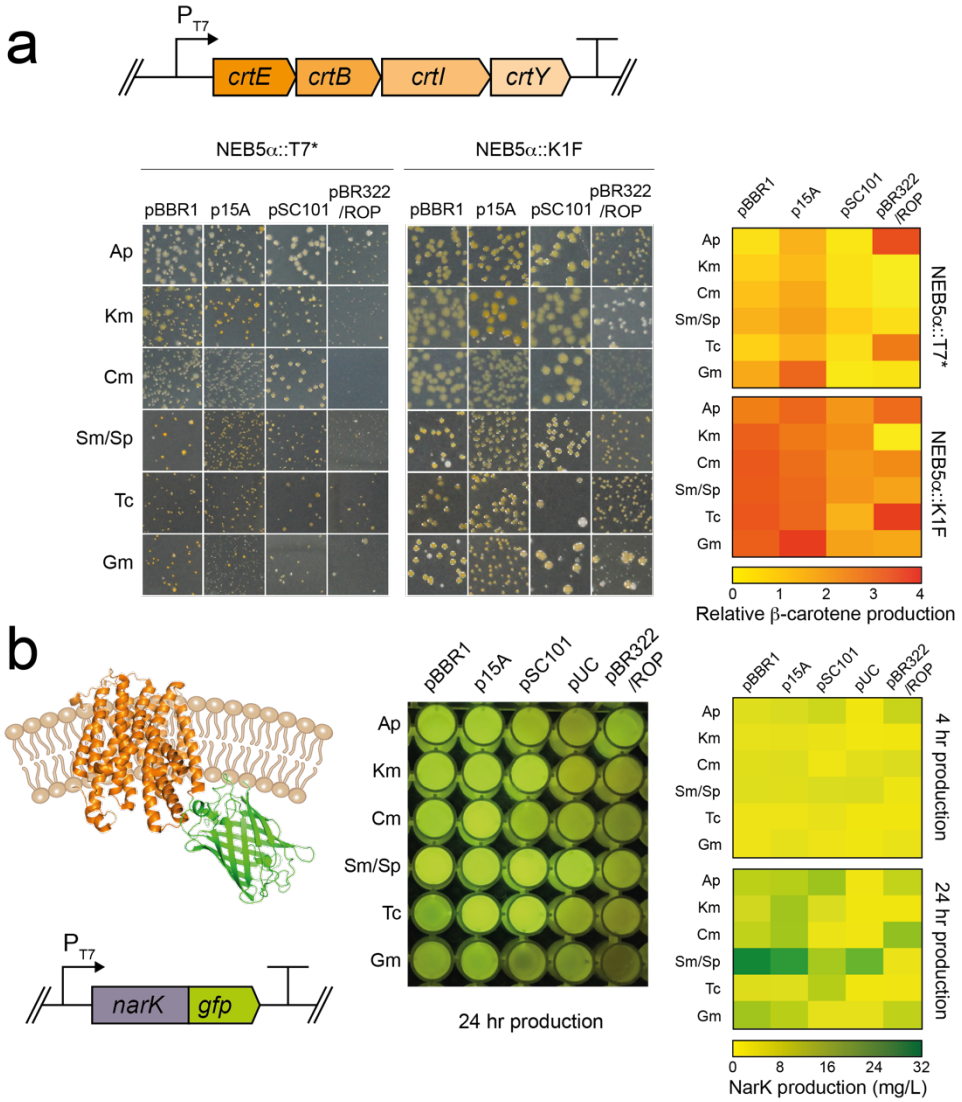


Figure 2. Production of β-carotene and the membrane protein NarK with 30 different combinations of origins of replication and antibiotic resistance markers. (a) Upper panel: illustration of the cargo with T7 promoter driven expression of the four-gene crtEBIY biosynthetic pathway. Lower left panel: colony phenotypes of the crtEBIY cargo combined with ampicillin (Ap), kanamycin (Km), chloramphenicol (Cm), spectinomycin/ streptomycin (Sm/Sp), tetracycline (Tc) or gentamycin selection; and the pBBR1, p15A, pSC101 or pBR322/ROP origins of replication. All 30 combinations were transformed into both *E. coli* NEB5α::T7* (left side) and NEB5α::K1F (right side) hosting two different variants of the T7 RNA polymerase. No combinations with the high copy pUC origin of replication yielded surviving colonies. Lower right panel: heat map representation of carotenoid levels measured by acetone extraction and absorbance at 453 nm on the 24 viable backbones variants in NEB5α::T7* and NEB5α::K1F. (b) Left panel: illustration of the NarK-GFP protein (based on the pdb files: 1EMA and 4U4V) and cargo constructs. Middle panel: Fluorescence from *E. coli* BL21 (DE3) transformed with 30 different backbones in combination with the T7-narK-gfp cargo. Right panel: heat map representation of fluorescence levels quantified in a microplate reader after four and 24 h expression.

References

(1) Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H., and Kompala, D. S. (1990) Plasmid-encoded protein: The principal factor in the "metabolic burden" associated with recombinant bacteria. Biotechnol. Bioeng. 35, 668−681.

(2) Grabherr, R., Nilsson, E., Striedner, G., and Bayer, K. (2002) Stabilizing plasmid copy number to improve recombinant protein production. Biotechnol. Bioeng. 77, 142−147.

(3) Friehs, K. (2004) Plasmid copy number and plasmid stability. In New Trends and Developments in Biochemical Engineering, pp 47−82, Springer, Berlin.

(4) Corchero, J. L., and Villaverde, A. (1998) Plasmid maintenance in Escherichia coli recombinant cultures is dramatically, steadily, and specifically influenced by features of the encoded proteins. Biotechnol. Bioeng. 58, 625−632.

(5) Ceroni, F., Algar, R., Stan, G. B., and Ellis, T. (2015) Quantifying cellular capacity identifies gene expression designs with reduced burden. Nat. Methods 12, 415−418.

(6) Weisse, A. Y., Oyarzun, D. A., Danos, V., and Swain, P. S. (2015) Mechanistic links between cellular trade-offs, gene expression, and growth. Proc. Natl. Acad. Sci. U. S. A. 112, E1038−1047.

(7) Kohanski, M. a, Dwyer, D. J., and Collins, J. J. (2010) How antibiotics kill bacteria: from targets to networks. Nat. Rev. Microbiol. 8, 423−435.

(8) Andersson, D. I., and Levin, B. R. (1999) The biological cost of antimicrobial resistance. Curr. Opin. Microbiol. 2, 489−493.

(9) Martínez, J. L., and Rojo, F. (2011) Metabolic regulation of antibiotic resistance. FEMS Microbiol. Rev. 35, 768−789.

(10) Schulz zur Wiesch, P., Engelstadter, J., and Bonhoeffer, S. (2010) Compensation of fitness costs and reversibility of antibiotic resistance mutations. Antimicrob. Agents Chemother. 54, 2085−2095.

(11) Levin, B. R., and Rozen, D. E. (2006) Non-inherited antibiotic resistance. Nat. Rev. Microbiol. 4, 556−562.

(12) Chao, R., Yuan, Y., and Zhao, H. (2014) Recent advances in DNA assembly technologies. FEMS Yeast Res. 15, 1−9.

(13) Cobb, R. E., Ning, J. C., and Zhao, H. (2014) DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. J. Ind. Microbiol. Biotechnol. 41, 469−477.

(14) Keasling, J. (2008) Synthetic biology for synthetic chemistry. ACS Chem. Biol. 3, 64−76.

(15) Vilanova, C., and Porcar, M. (2014) iGEM 2.0−refoundations for engineering biology. Nat. Biotechnol. 32, 420−424.

(16) Casini, A., Storch, M., Baldwin, G. S., and Ellis, T. (2015) Bricks and blueprints: methods and standards for DNA assembly. Nat. Rev. Mol. Cell Biol. 16, 568−576.

(17) Vilanova, C., Tanner, K., Dorado-Morales, P., Villaescusa, P., Chugani, D., Frías, A., Segredo, E., Molero, X., Fritschi, M., Morales, L., Ramon, D., Pena, C., Pereto, J., and Porcar, M. (2015) Standards´not that standard. J. Biol. Eng. 9, 17.

(18) Silva-Rocha, R., Martínez-García, E., Calles, B., Chavarría, M., Arce-Rodríguez, A., Heras, A. D. L., Paez-Espino, A. D., Durante-Rodríguez, G., Kim, J., Nikel, P. I., Platero, R., and de Lorenzo, V. (2013) The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. Nucleic Acids Res. 41, 666−675.

(19) Martínez-García, E., Aparicio, T., Goñi-Moreno, A., Fraile, S., and de Lorenzo, V. (2015) SEVA 2.0: an update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. Nucleic Acids Res. 43, D1183−D1189.

(20) Gibson, D. G., Benders, G. A., Axelrod, K. C., Zaveri, J., Algire, M. A., Moodie, M., Montague, M. G., Venter, J. C., Smith, H. O., and Hutchison, C. A. (2008) One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. Proc. Natl. Acad. Sci. U. S. A. 105, 20404−20409.

(21) Nour-eldin, H. H., Hansen, B. G., Nørholm, M. H. H., Jensen, J. K., and Halkier, B. A. (2006) Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. Nucleic Acids Res. 34, e122.

(22) Engler, C., Kandzia, R., and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. PLoS One 3, e3647.
(23) Shetty, R. P., Endy, D., and Knight, T. F. J. (2008) Engineering BioBrick vectors from BioBrick parts. J. Biol. Eng. 2, 5.
(24) Cavaleiro, A. M., Kim, S. H., Seppalä , S., Nielsen, M. T., and Nørholm, M. H. H. (2015) Accurate DNA assembly and genome engineering with optimized uracil excision cloning. ACS Synth. Biol. 4,
1042−1046.
(25) Cavaleiro, A. M., Nielsen, M. T., Kim, S. H., Seppalä, S., and Nørholm, M. H. H. (2015) Uracil excision for assembly of complex pathways. In Springer Protocol Handbooks, pp 1−11, Humana Press,
Springer-Verlag, Berlin.
(26) Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F., and Voigt, C. A. (2012) Modular control of multiple pathways using engineered orthogonal T7 polymerases. Nucleic Acids Res. 40, 8773−8781.
(27) Mirzadeh, K., Martínez, V., Toddo, S., Guntur, S., Herrgard, M. J., Elofsson, A., Nørholm, M. H. H., and Daley, D. O. (2015) Enhanced protein production in Escherichia coli by optimization of cloning scars at the vector−coding sequence junction. ACS Synth. Biol. 4, 959−965.

Supporting Information for
SEVA Linkers: A Versatile and Automatable DNA Backbone Exchange Standard for Synthetic Biology

Materials and Methods

Strains, media and plasmids

Escherichia coli NEB5α (New England Biolabs, Ipswich, MA, USA) was used for propagation of plasmids, backbone swapping optimization and as a general cloning host except when *E. coli* DB3.1 (Thermo Fisher Scientific, Invitrogen, Waltham, MA, USA) was used for handling of *ccdB*-containing plasmids. NEB5α::T7* and NEB5α::K1F[1] were used for β-carotene production. SOC media was used as a recovery media after transformation. Bacteria were propagated in Luria-Bertani (LB) liquid media or agar plates supplemented with ampicillin (100 μg/ml), kanamycin (50 μg/ml), chloramphenicol (50 μg/ml), spectinomycin (50 μg/ml), tetracycline (10 μg/ml), or gentamycin (10 μg/ml) when required. In most cases LB media was used for liquid cultures except for carotenoid production assessment that was accomplished in 2×YT media supplemented with 0.5% glycerol.

Molecular biology reagents

T4 DNA ligase and restriction enzymes were purchased from Thermo Fischer Scientific (Waltham, MA, USA). Nicking enzymes were purchased from New England Biolabs (Ipswich, MA, USA). Plasmids were isolated using the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany). PCR

products were purified using a PureLinkTM Quick Gel Extraction and PCR Purification Combo Kit (Life Technologies, Foster City, USA). Buffers for PCR and cloning reactions were purchased from Agilent Technologies (Santa Clara, CA, USA) and Thermo Fischer Scientific (Wilmington, USA). PCR was performed with the proofreading PfuX7 polymerase as previously described.[3]

Plasmid constructions

SEVA linker sequences (version #1, see Supporting Information Figure S3) were introduced into pACYCDuet-1 and pCDFDuet-1 (for references to plasmids see Supporting Information Table S2) by amplifying the backbones with oligonucleotides #1 and #2 (for numbering and sequences of oligonucleotides see Supporting Information Table S1), gfp from pETDuet-1-gfp with the oligonucleotides #3 and #4 and lacZα from pBluescript II KS (+) with oligonucleotides #5 and #6, followed by assembly by uracil excision as described previously.1, 2 This created pACYC-sl1- gfp and pCDF-sl1-gfp.

The ccdB gene was amplified from pOSIP-KT using oligonucleotides #7 and #8 and mixed with a pACYC-sl1 fragment obtained from Nb.BtsI-digested pACYC- sl1-gfp creating pACYC-sl1-ccdB upon transformation into *E. coli*.

pCDF-sl2-gfp was constructed by amplifying the vector backbone and gfp insert from pCDF-sl1-gfp with oligonucleotides #9 and #10, and #11 and #12, respectively, treating the PCR products with Nb.BtsI and transforming the fragments into *E. coli*. pACYC-sl2-ccdB was similarly obtained by combining Nb.BtsI-digested pACYC-sl1-ccdB with a backbone fragment amplified from pACYC-sl1-ccdB with oligonucleotides #9 and #10.

pCDF-sl3-gfp and pACYC-sl3-ccdB were created by amplifying the corresponding sl2 versions with oligonucleotides #13 and #14, gfp with #15 and #16, and ccdB with #17 and #18, followed by Nt.BbvCI-treatment and transformation.

pCDF-sl4-gfp and pACYC-sl4-ccdB were created by amplifying the corresponding sl2 versions with the oligonucleotides #19 and #20, gfp with #21 and #22, and ccdB with #23 and #24 followed by uracil excision cloning.

pSEVAXX-sl3-ccdB series was constructed by combining five origins of replication (pBBR1; #3, p15A; #6, pSC101; #7, pUC; #8, pBR322/ROP; #9), prepared by digestion with PacI and PshAI from the corresponding parts in the pSEVA collection, and six antibiotic resistance markers (ampicillin; #1, kanamycin; #2, chloramphenicol; #3, spectinomycin; #4, tetracycline; #5, gentamycin; #6), prepared by digestion with PshAI and SpeI from the corresponding parts in the pSEVA collection, with the sl3-ccdB cargo isolated after digestion by PacI and SpeI from pACYC-sl3-ccdB.

The β-carotene biosynthetic pathway was introduced into pACYC-sl3 by SEVA cloning (see below) after amplifying the crtEBIY operon from pSIJ31B with the oligonucleotides #25 and #26, creating pACYC-sl3-T7-crtEBIY. Combining pACYC-sl3-T7-crtEBIY with the pSEVAXX-sl3-ccdB series as described below created the pSEVAXX-sl3-T7-crtEBIY series.

The T7-narK-gfp cargo was amplified from pET28a-narKWTP7-6 using oligonucleotides #29 and #30, and mixed with a pSEVAXX-sl3 backbone amplified with oligonucleotides #27 and #28 followed by uracil excision cloning. Combining the initial pSEVAXX-sl3-T7-narK-gfp clone with the pSEVAXX-sl3-ccdB series created the complete pSEVAXX-sl3- T7-narK-gfp series.

Nicking enzyme mediated one-tube backbone exchange (SEVA cloning)

0.06 pmol of each plasmid were added to a 10 or 20 µl total reaction volume containing CutSmart® buffer (New England Biolabs, Ipswich, MA, USA) and five units of Nt.BbvCI. Mixtures were kept at 37°C for 1 h, followed by 25°C for 15 min, 10°C for 10 min, then incubation at 0°C using a C1000 TouchTM Thermal Cycler (Bio-Rad, Hercules, CA, USA). 2.5 U of T4 DNA ligase and buffer were added, followed by incubation at room temperature for 15 min and storage on ice prior to transformation.

Gibson assembly- and uracil excision-based backbone exchange

The GFP cargo was PCR amplified from pCDF-sl3-gfp using the oligonucleotides #31 and #32 for Gibson assembly and #35 and #36 for uracil excision; and the backbone was amplified from pACYC-sl3-ccdB using the oligonucleotides #33 and #34 (Gibson) or #37 and #38 (uracil excision). After PCR, DpnI treatment was done at 37°C for 40 min followed by gel purification. Insert and vector ratio of 2:1 was applied for both methods. Gibson assembly was performed with 2X Gibson Assembly Master Mix (New England Biolabs, Ipswich, MA, USA) following the manufacturer´s instructions. Assembly by uracil excision was as described previously.[1, 2]

Production and relative quantification of β-carotene

The NEB5αT7* and NEB5αK1F strains1 were used to compare β-carotene productivity in the pSEVA-crtEBIY series. Corresponding strains with the pathway integrated in one copy on the genome (NEB5αT7*::EBIY and NEB5αK1F::EBIY) were used as reference strains.1 Cells were grown in 2×YT media supplemented with 0.5% glycerol at 30°C for 72 hours with 300 rpm. 1 ml of each culture was harvested by centrifugation at 13000 rpm for 5 min. After discarding the supernatant, cells were washed once with 1 ml of water. 1 ml of acetone was added and the pellets

re- suspended vigorously by vortexing, followed by incubation at 55°C for 20 min at 1000 rpm using a tabletop shaker. The remaining cell debris were removed by centrifugation (13000 rpm, 5 min) and absorbance at 453 nm was measured in a UV- 1600PC spectrophotometer (VWR International, Radnor, PA, USA) on 500 µl extract using a quartz cuvette. The equation for calculating β-carotene production titers is $\frac{A\ x\ Vol(mL)\ x\ 10^4}{E_{1\ cm}^{1\ \%}}$ where A is the UV/Vis absorbance at 453 nm and $E_{1\ cm}^{1\ \%}$ is the specific absorption coefficient ($E_{1\ cm}^{1\ \%}$, β-carotene=2503).

Production and quantification of NarK-GFP

The pSEVAXX-sl3- T7-narK-gfp series was transformed into BL21 (DE3) (Novagen, Madison, WI, USA) using a standard protocol. Overnight cultures were prepared by inoculating a single colony in 800 µL of LB liquid media containing the different pSEVA-defined antibiotics (100 µg/ml ampicillin; 50 µg/ml kanamycin; 34 µg/ml chloramphenicol; 50 µg/ml spectinomycin; 10 µg/ml tetracycline; 10 µg/ml gentamycin) in 96-deep well plates at 37 °C and 300 rpm in an Innova 44 incubator (Thermo Scientific, Waltham, MA, USA). For expression measurements, overnight cultures were back-diluted 1:50 in 3 ml LB media containing the different antibiotics in 24-well plates at 37 °C and 300 rpm. When exponential phase (an OD600 of approximately 0.5) was reached, expression was induced by addition of 1 mM IPTG and incubated at 25 °C and 300 rpm for 2 hours. Subsequently, 1 ml of culture was harvested at 2270 x g for 20 min, resuspended in a buffer (50 mM TrisŊHCl (pH 8.0), 200 mM NaCl, and 15 mM EDTA) and incubated for 2 hours at room temperature. Fluorescence was measured in a 96-well plate in a SynergyTM Mx plate reader (BioTek) (excitation wavelength 485 nm; emission wavelength 512 nm). The amount of protein produced was estimated from a GFP standard curve. The standard curve was obtained from purified GFP mixed with BL21 (DE3) cells to account for quenching effects. 22 h after induction the remaining culture was harvested, resuspended in buffer, incubated for 2 hours and fluorescence measured as described above.

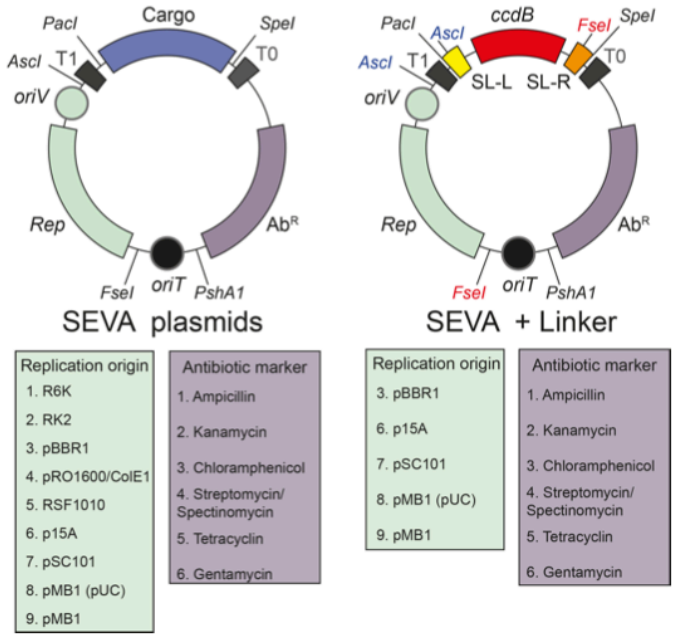


Figure S1. Illustration of the Standard European Vector Architecture (SEVA) system and the relationship to the SEVA linkers and plasmids described here. Left panel: In the original SEVA system, rare restriction sites flank the three basic components: origins of replication, antibiotic selection markers and the cargo. This basic design enables exchange of the basic components using classical restriction enzyme molecular cloning. Right panel: The new SEVA linkers enable simple, one-pot backbone shuffling by introducing two multifunctional linker sequences (highlighted in yellow and orange color) flanking the cargo in the SEVA system. In this example, the ccdB counterselection marker is the cargo. Any cargo that is flanked by SEVA linkers can be converted to the SEVA system by utilizing the PacI and SpeI restriction sites, but also by a range of other molecular cloning technologies (see main text). The SEVA linkers violate the basic SEVA design rules by reusing the AscI and FseI sites (highlighted in blue and red font).

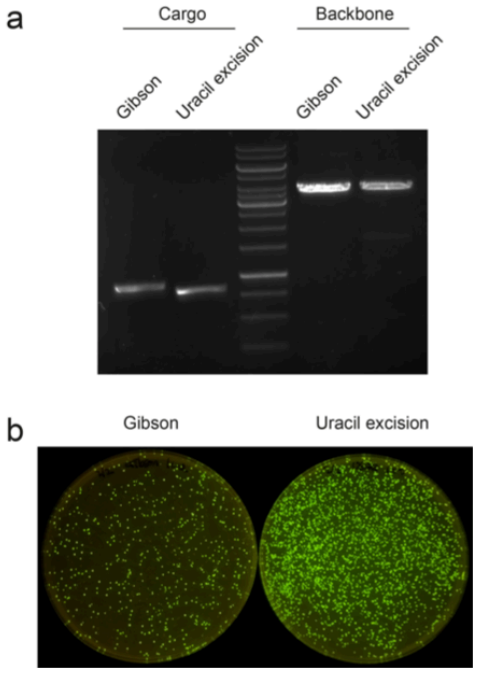


Figure S2. Demonstration of amplification and reassembly of SEVA linker flanked cargo and backbone with Gibson assembly and uracil excision cloning. (a) Agarose gel showing PCR amplified cargoes and backbones compatible with Gibson assembly or uracil excision cloning. In the cargo, SEVA linkers flanked a transcriptional unit (a leaky Ptrc promoter driving expression of gfp). For details on the protocol see Materials and Methods. (b) PCR-amplified cargo and backbone was reassembled with Gibson and uracil excision cloning and plated on agar plates supplemented with the antibiotic corresponding to the backbone selection marker. Green fluorescent colonies demonstrate the presence of the cargo.

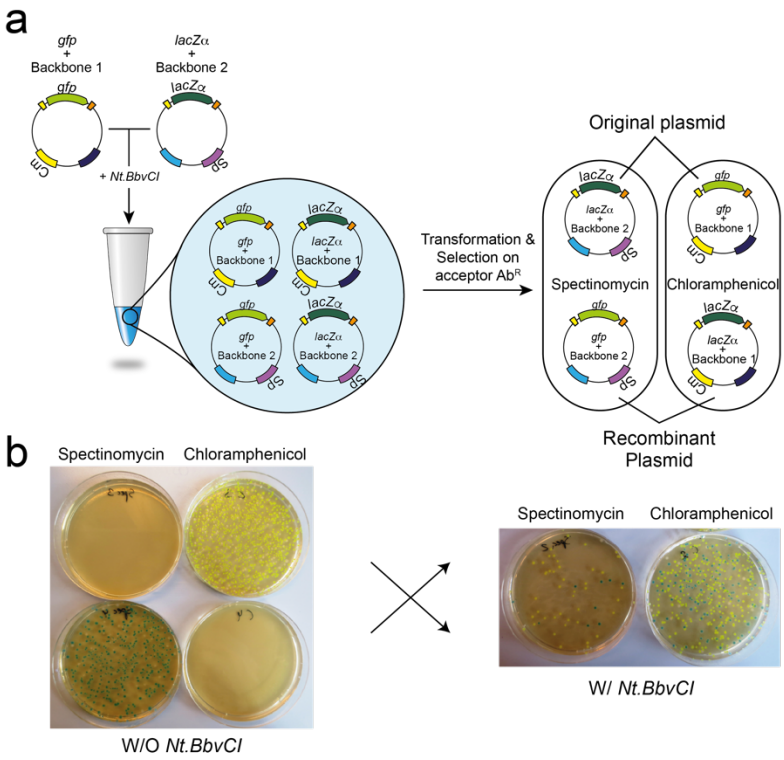


Figure S3. Exchange of DNA fragments using four Nt.BbvCI- nicking enzyme- sites in the SEVA linkers. Yellow and orange boxes mark the four Nt.BbvCI recognition sites. The two three-nucleotide spacers between the two double nicking sites are different from each other, thereby ensuring specificity and directionality in the parts exchange. Any cargo (light blue box) of choice can replace the ccdB (red box) counterselection marker.

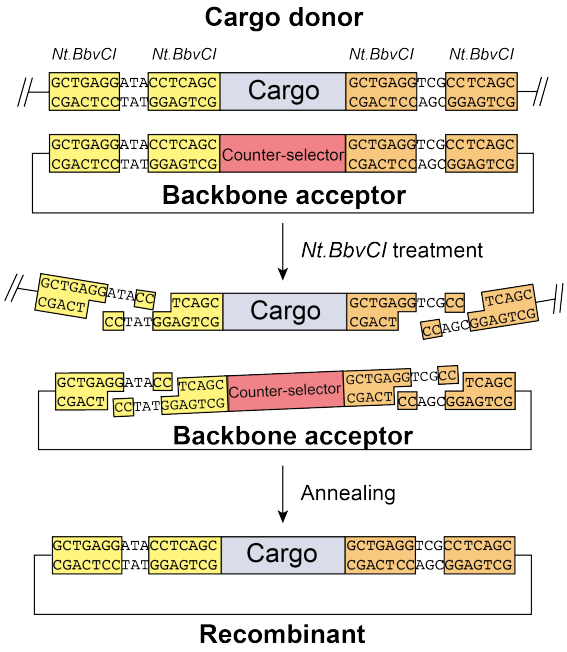


Figure S4. Demonstration of parts exchange between two plasmids that contain SEVA linkers. (a) Illustration of a gfp expressing SEVA linker cargo in a plasmid that confers chloramphenicol resistance, mixed with a lacZα expressing SEVA linker cargo in a plasmid conferring spectinomycin resistance, leading to four different recombinant plasmids. (b) Left panel: The two different plasmids were transformed into NEB5α and plated on LB agar supplemented with X-gal combined with spectinomycin or chloramphenicol. Right panel: The two plasmids were mixed with the nicking enzyme Nb.BtsI, transformed into NEB5α and plated on LB X-gal agar.

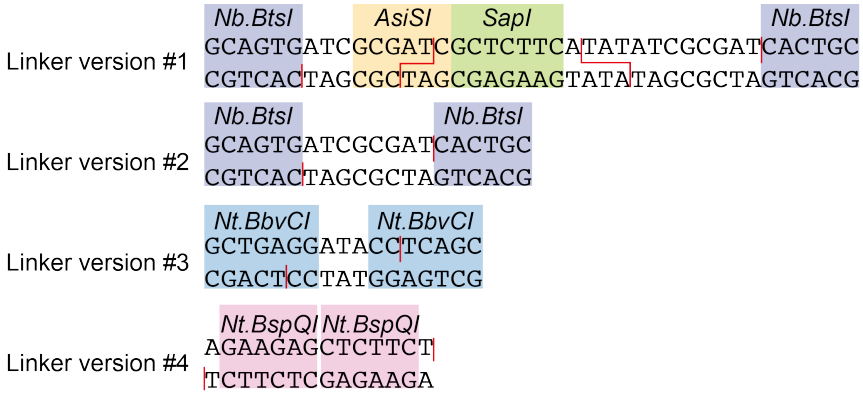


Figure S5. Different types of nicking enzyme-based linkers that were tested for cloning efficiency and accuracy. Version #2 outperforms version #1 in simplicity, but the Nb.BtsI-based designs generally performed poorly, probably due to the frequent occurrence of the six-nucleotide recognition sites outside the SEVA linkers. Version #4 likewise performed poorly, probably because the single-strand overhangs formed by the Nt.BspQI sites were too stable. Version #3 clearly outperformed the other designs in terms of cloning efficiency and accuracy and was the preferred choice for the SEVA linkers.

Table S1. Oligonucleotides used in this study

No.	NAME	SEQUENCE
1	Duet-sl1-rev	ATCGCGAUCACTGCCGCGCGCGTCGACGGGCGCGCCGCGGCCGCTTAATTAACAAA ATTATTTCTACAGGGGAATTGTTATCCGCTC
2	Duet-sl1-fwd	AGAGCGAUCGCACTCACTGCGGCCGGCCCCTGCAGGCGCCGGCGACTAGTCCTAGGCT GCTGCCACCGCTG
3	Sl1-PtrcGFP_fwd	ATCGCGAUCGCTCTTCATATATCGCGATCACTGCTTGACAATTAATCATCCGGCTCGTA TAATG
4	Sl1-PtrcGFP_rev	ATCGCTCUTCATATATCGCGATCACTGCTTATTTGTAGAGCTCATCCATGCCATGTG
5	Sl1-lacZa_fwd	ATCGCGAUCGCTCTTCATATATCGCGATCACTGCACCAGTGGNTCATCTCCAAGCAGTG GTTCGCGCAACGCAATTAATGTGAG
6	Sl1-lacZa_rev	AGTGCGAUCGCTCTTCATATAGTGCGATCACTGCACCAGTGCTACCTCCTGAACCAC
7	Sl1-ccdB_fwd	ATC GCG AUC GCT CTT CAT ATA TCG CGA TCA CTG CTA CTA AAA GCC AGA TAA CAG TAT GCG TAT
8	Sl1-ccdB_rev	AGT GCG AUC GCT CTT CAT ATA GTG CGA TCA CTG CCG GGT TAT TAT ATT CCC CAG AAC ATC AG
9	Nb.BtsI_SL_F (vector)	ATCGCactCACTGCGGCCGGCCCCTG
10	Nb.BtsI_SL_R (vector)	ATCGCgatCACTGCCGCGCGCGCGTC
11	Nb.BtsI_PtrcGFP_F	ATCGCGATCACTGCTTGACAATTAATCATCCGGCTCGTATAATG
12	Nb.BtsI_PtrcGFP_R	AGTGCGATCACTGCTTATTTGTAGAGCTCATCCATGCCATGTG
13	Nt.BbvCI_V_F	GGTCGCCTCAGCGGCCGGCCCCTGCAGGCG
14	Nt.BbvCI_V_R	GGTATCCTCAGCCGCGCGCGCGTCGACGGG
15	Nt.BbvCI_GFP_F	GGATACCTCAGCTTGACAATTAATCATCCGGC
16	Nt.BbvCI_GFP_R	GGCGACCTCAGCTTATTTGTAGAGCTCATCCATGC
17	Nt.BbvCI_CCDB_F	GGATACCTCAGCTACTAAAAGCCAGATAACAGTATGC
18	Nt.BbvCI_CCDB_R	GGCGACCTCAGCCGGGTTATTATATTCCCCAG
19	Nt.BspQ1_V_R_U	AGA AGA GCT CTT CUC GCG CGC GCG TCG ACG GG
20	Nt.BspQ1_V_F_U	AGAAGAGCGCTCTTCUGGCCGGCCCCTGCAGGCG
21	Nt.BspQ1_GFP_F_U	AGAAGAGCTCTTCUTTGACAATTAATCATCCGGC
22	Nt.BspQ1_GFP_R_U	AGA AGA GCG CTC TTC UTT ATT TGT AGA GCT CAT CCA TGC
23	Nt.BspQ1_ccdB_F_U	AGAAGAGCTCTTCUtactaaaagccagataacagtatgc
24	Nt.BspQ1_ccdB_R_U	AGA AGA GCG CTC TTC UCG GGT TAT TAT ATT CCC CAG AAC ATC AG
25	Nt.BbvCI_EBIY_F	GGATACCTCAGCGGATCTCGACGCTCTCCC
26	Nt.BbvCI_EBIY_R	GGCGACCTCAGCGATTATGCGG
27	SL3-backbone_rev	AGGTATCCUCAGCCGCGCG
28	SL3-backbone_fwd	AGCTGAGGUCGCCTCAGC
29	NarK-sl3_fwd	AGGATACCUCAGCTAATACGACTCACTATAGGG
30	NarK-sl3_rev	ACCTCAGCUCAGTGGTGGTGG
31	SL-Gibson_ifwd	TTAATTAAGCGGCCGCGGCGCGCCCGTCGA
32	SL-Gibson_irev	ACTAGTCGCCGGCGCCTGCAGGGGCCGGCC
33	SL-Gibson_bbfwd	GGCCGGCCCCTGCAGGCGCCGGCGACTAGT
34	SL-Gibson_bbrev	TCGACGGGCGCGCCGCGGCCGCTTAATTAA
35	SL3m-Ptrc_U_F	AGGATACCUCAGCTTGACAATTAATC
36	SL3m-GFP_U_R	ACCTCAGCUTATTTGTAGAGCTC
37	pCDF_BB_U_F	AGCTGAGGUCGCCTCAGC
38	pCDF_BB_U_R	AGGTATCCUCAGCCGCGCG

Table S2. Strains and plasmids used in this study

Strain/plasmid	Property	Source/Reference
Strains		
E. coli NEB5α	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	NEB
E. coli DB3.1	F- gyrA462 endA1 glnV44 Δ(sr1-recA) mcrB mrr hsdS20(r_B^-, m_B^-) ara14 galK2 lacY1 proA2 rpsL20(Sm^r) xyl5 Δleu mtl1	Thermo Fisher Scientific
NEB5α::T7*	NEB5α with a T7* RNA polymerase integrated	(1)
NEB5α::K1F	NEB5α with a T7*(K1F) RNA polymerase integrated	(1)
BL21 (DE3)	F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Novagen
Plasmids		
pCDFDuet-1	Cloning and expression vector, Sp^R	Novagen
pACYCDuet-1	Cloning and expression vector, Cm^R	Novagen
pETDuet-1-*gfp*	Constitutively expressed GFP, Amp^R	(1)
pBluescript II KS(+)	*lacZ* ✔ expressed from lac promoter	Agilent Technology
pCDF-sl1- *lacZ* ✔	*lacZα* fragment flanked by SEVA linker ver1, Sp^R, CloDF13 origin from pCDFDuet-1	This study
pACYC-sl1-*gfp*	*gfp* flanked by SEVA linker ver1, Cm^R, p15A origin from pACYCDuet-1	This study
pCDF-sl2-*gfp*	*gfp* flanked by SEVA linker ver2, Sp^R, CloDF13 origin from pCDFDuet-1	This study
pACYC-sl2-*ccdB*	*ccdB* flanked by SEVA linker ver2, Cm^R, p15A origin from pACYCDuet-1	This study
pCDF-sl3- *gfp*	*gfp* flanked by SEVA linker ver3, Sp^R, CloDF13 origin from pCDFDuet-1	This study
pACYC-sl3-*ccdB*	*ccdB* flanked by SEVA linker ver3, Cm^R, p15A origin from pACYCDuet-1	This study
pCDF-sl4- *gfp*	*gfp* flanked by SEVA linker ver4, Sp^R, CloDF13 origin from pCDFDuet-1	This study
pACYC-sl4-*ccdB*	*ccdB* flanked by SEVA linker ver4, Cm^R, p15A origin from pACYCDuet-1	This study
pSIJ31B	*P. ananatis crtEBIY* operon, Sp^R, CloDF13 origin	Unpublished
pOSIP-KT	P21 Integration module, *ccdB*, Km^R, pUC origin	(5)
pET28a-*narK*^{WT}P7-6 P7-6	*narK*-gfp with T7 promoter, Km^R, pBR322/ROP origin	(6)
pSEVA13-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Amp^R, pBBR1 origin	This study
pSEVA16-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Amp^R, p15A origin	This study
pSEVA17-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Amp^R, pSC101 origin	This study
pSEVA18-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Amp^R, pUC origin	This study
pSEVA19-sl3-*ccdB*	*ccdB* flanked by SEVAlinkers, Amp^R, pBR322/ROP origin	This study
pSEVA23-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Km^R, pBBR1 origin	This study
pSEVA26-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Km^R, p15A origin	This study
pSEVA27-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Km^R, pSC101 origin	This study
pSEVA28-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Km^R, pUC origin	This study
pSEVA29-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Km^R, pBR322/ROP origin	This study
pSEVA33-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Cm^R, pBBR1 origin	This study
pSEVA36-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Cm^R, p15A origin	This study
pSEVA37-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Cm^R, pSC101 origin	This study
pSEVA38-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Cm^R, pUC origin	This study

pSEVA39-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Cm^R, pBR322/ROP origin	This study
pSEVA43-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Sp^R, pBBR1 origin	This study
pSEVA46-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Sp^R, p15A origin	This study
pSEVA47-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Sp^R, pSC101 origin	This study
pSEVA48-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Sp^R, pUC origin	This study
pSEVA49-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Sp^R, pBR322/ROP origin	This study
pSEVA53-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Tet^R, pBBR1 origin	This study
pSEVA56-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Tet^R, p15A origin	This study
pSEVA57-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Tet^R, pSC101 origin	This study
pSEVA58-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Tet^R, pUC origin	This study
pSEVA59-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Tet^R, pBR322/ROP origin	This study
pSEVA63-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Gm^R, pBBR1 origin	This study
pSEVA66-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Gm^R, p15A origin	This study
pSEVA67-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Gm^R, pSC101 origin	This study
pSEVA68-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Gm^R, pUC origin	This study
pSEVA69-sl3-*ccdB*	*ccdB* flanked by SEVA linkers, Gm^R, pBR322/ROP origin	This study
pSEVA13-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Amp^R, pBBR1 origin	This study
pSEVA16-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Amp^R, p15A origin	This study
pSEVA17-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Amp^R, pSC101 origin	This study
pSEVA19-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Amp^R, pBR322/ROP origin	This study
pSEVA23-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Km^R, pBBR1 origin	This study
pSEVA26-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Km^R, p15A origin	This study
pSEVA27-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Km^R, pSC101 origin	This study
pSEVA29-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Km^R, pBR322/ROP origin	This study
pSEVA33-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Cm^R, pBBR1 origin	This study
pSEVA36-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Cm^R, p15A origin	This study
pSEVA37-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Cm^R, pSC101 origin	This study
pSEVA39-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Cm^R, pBR322/ROP origin	This study
pSEVA43-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Sp^R, pBBR1 origin	This study
pSEVA46-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Sp^R, p15A origin	This study
pSEVA47-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Sp^R, pSC101 origin	This study
pSEVA49-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Sp^R, pBR322/ROP origin	This study
pSEVA53-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Tet^R, pBBR1 origin	This study

pSEVA56-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Tet^R, p15A origin	This study
pSEVA57-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Tet^R, pSC101 origin	This study
pSEVA59-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Tet^R, pBR322/ROP origin	This study
pSEVA63-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Gm^R, pBBR1 origin	This study
pSEVA66-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Gm^R, p15A origin	This study
pSEVA67-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Gm^R, pSC101 origin	This study
pSEVA69-sl3-T7-*crtEBIY*	*crtEBIY* operon flanked by SEVA linkers, Gm^R, pBR322/ROP origin	This study
pSEVA13NarK/GFP	*narK-gfp* flanked by SEVA linkers, Amp^R, pBBR1 origin	This study
pSEVA16-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Amp^R, p15A origin	This study
pSEVA17-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Amp^R, pSC101 origin	This study
pSEVA18-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Amp^R, pUC origin	This study
pSEVA19-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Amp^R, pBR322/ROP origin	This study
pSEVA23-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Km^R, pBBR1 origin	This study
pSEVA26-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Km^R, p15A origin	This study
pSEVA27-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Km^R, pSC101 origin	This study
pSEVA28-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Km^R, pUC origin	This study
pSEVA29-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Km^R, pBR322/ROP origin	This study
pSEVA33-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Cm^R, pBBR1 origin	This study
pSEVA36-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Cm^R, p15A origin	This study
pSEVA37-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Cm^R, pSC101 origin	This study
pSEVA38-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Cm^R, pUC origin	This study
pSEVA39-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Cm^R, pBR322/ROP origin	This study
pSEVA43-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Sp^R, pBBR1 origin	This study
pSEVA46-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Sp^R, p15A origin	This study
pSEVA47-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Sp^R, pSC101 origin	This study
pSEVA48-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Sp^R, pUC origin	This study
pSEVA49-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Sp^R, pBR322/ROP origin	This study
pSEVA53 -sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Tet^R, pBBR1 origin	This study
pSEVA56 -sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Tet^R, p15A origin	This study
pSEVA57-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Tet^R, pSC101 origin	This study
pSEVA58-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Tet^R, pUC origin	This study
pSEVA59-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Tet^R, pBR322/ROP origin	This study
pSEVA63-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Gm^R, pBBR1 origin	This study

pSEVA66-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Gm^R, p15A origin	This study
pSEVA67-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Gm^R, pSC101 origin	This study
pSEVA68-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Gm^R, pUC origin	This study
pSEVA69-sl3- T7-*narK-gfp*	*narK-gfp* flanked by SEVA linkers, Gm^R, pBR322/ROP origin	This study

Table S3. β-carotene and NarK production overview

Strain/pSEVA construct	Resistance/origin	Production (mg/L)
NEB5α::T7*	/genome	1.29
NEB5α::T7*::Km^R	Km/genome	1.30
13	Amp/pBBR1	0.64
16	Amp/p15A	1.56
17	Amp/pSC101	0.54
19	Amp/pBR322/ROP	3.69
23	Km/pBBR1	0.97
26	Km/p15A	1.37
27	Km/pSC101	0.61
29	Km/pBR322/ROP	0.38
33	Cm/pBBR1	1.24
36	Cm/p15A	1.70
37	Cm/pSC101	0.63
39	Cm/pBR322/ROP	0.31
43	Sp/pBBR1	1.54
46	Sp/p15A	1.91
47	Sp/pSC101	1.03
49	Sp/pBR322/ROP	0.67
53	Tet/pBBR1	0.95
56	Tet/p15A	1.51
57	Tet/pSC101	0.66
59	Tet/pBR322/ROP	2.69
63	Gm/pBBR1	1.72
66	Gm/p15A	3.16
67	Gm/pSC101	0.51
69	Gm/pBR322/ROP	0.59
NEB5α::K1F		
13	Amp/pBBR1	2.58
16	Amp/p15A	3.15
17	Amp/pSC101	2.08
19	Amp/pBR322/ROP	3.03
23	Km/pBBR1	3.28
26	Km/p15A	2.72
27	Km/pSC101	2.33
29	Km/pBR322/ROP	0.43
33	Cm/pBBR1	3.46
36	Cm/p15A	2.97
37	Cm/pSC101	2.07
39	Cm/pBR322/ROP	2.36
43	Sp/pBBR1	3.41
46	Sp/p15A	3.13
47	Sp/pSC101	2.11
49	Sp/pBR322/ROP	1.83
53	Tet/pBBR1	3.51
56	Tet/p15A	3.17
57	Tet/pSC101	1.53
59	Tet/pBR322/ROP	4.10
63	Gm/pBBR1	3.31
66	Gm/p15A	4.15
67	Gm/pSC101	1.85
69	Gm/pBR322/ROP	1.73
BL21 (DE3), 4hr		
13	Amp/pBBR1	9.32
16	Amp/p15A	11.84

17	Amp/pSC101	17.21
18	Amp/pUC	0.64
19	Amp/pBR322/ROP	19.43
23	Km/pBBR1	5.50
26	Km/p15A	6.26
27	Km/pSC101	3.49
28	Km/pUC	0.21
29	Km/pBR322/ROP	0.54
33	Cm/pBBR1	8.76
36	Cm/p15A	10.46
37	Cm/pSC101	1.70
38	Cm/pUC	7.13
39	Cm/pBR322/ROP	10.91
43	Sp/pBBR1	9.25
46	Sp/p15A	9.08
47	Sp/pSC101	11.34
48	Sp/pUC	11.79
49	Sp/pBR322/ROP	1.92
53	Tet/pBBR1	2.48
56	Tet/p15A	2.28
57	Tet/pSC101	4.72
58	Tet/pUC	0.55
59	Tet/pBR322/ROP	0.70
63	Gm/pBBR1	2.52
66	Gm/p15A	5.61
67	Gm/pSC101	2.37
68	Gm/pUC	1.39
69	Gm/pBR322/ROP	3.77
BL21 (DE3), 24hr		
13	Amp/pBBR1	24.29
16	Amp/p15A	26.68
17	Amp/pSC101	36.69
18	Amp/pUC	0.95
19	Amp/pBR322/ROP	21.08
23	Km/pBBR1	15.94
26	Km/p15A	35.15
27	Km/pSC101	11.36
28	Km/pUC	0.47
29	Km/pBR322/ROP	0.72
33	Cm/pBBR1	22.26
36	Cm/p15A	34.44
37	Cm/pSC101	2.71
38	Cm/pUC	0.88
39	Cm/pBR322/ROP	42.22
43	Sp/pBBR1	91.58
46	Sp/p15A	78.48
47	Sp/pSC101	32.66
48	Sp/pUC	57.18
49	Sp/pBR322/ROP	3.16
53	Tet/pBBR1	9.63
56	Tet/p15A	9.29
57	Tet/pSC101	27.03
58	Tet/pUC	0.75
59	Tet/pBR322/ROP	0.96
63	Gm/pBBR1	35.26
66	Gm/p15A	25.26
67	Gm/pSC101	5.80
68	Gm/pUC	5.96
69	Gm/pBR322/ROP	23.14

References

(1) Cavaleiro, A. M., Kim, S. H., Seppälä, S., Nielsen, M. T., and Nørholm, M. H. H. (2015) Accurate DNA assembly and genome engineering with optimized uracil excision cloning. ACS Synt. Biol. 4, 1042-1046.

(2) Cavaleiro, A. M., Nielsen, M. T., Kim, S. H., Seppälä, S., and Nørholm, M. H. H. (2015) Uracil excision for assembly of complex pathways. In Springer Protocol Handbooks, pp1−11, Humana Press, Springer−Verlag Berlin Heidelberg.

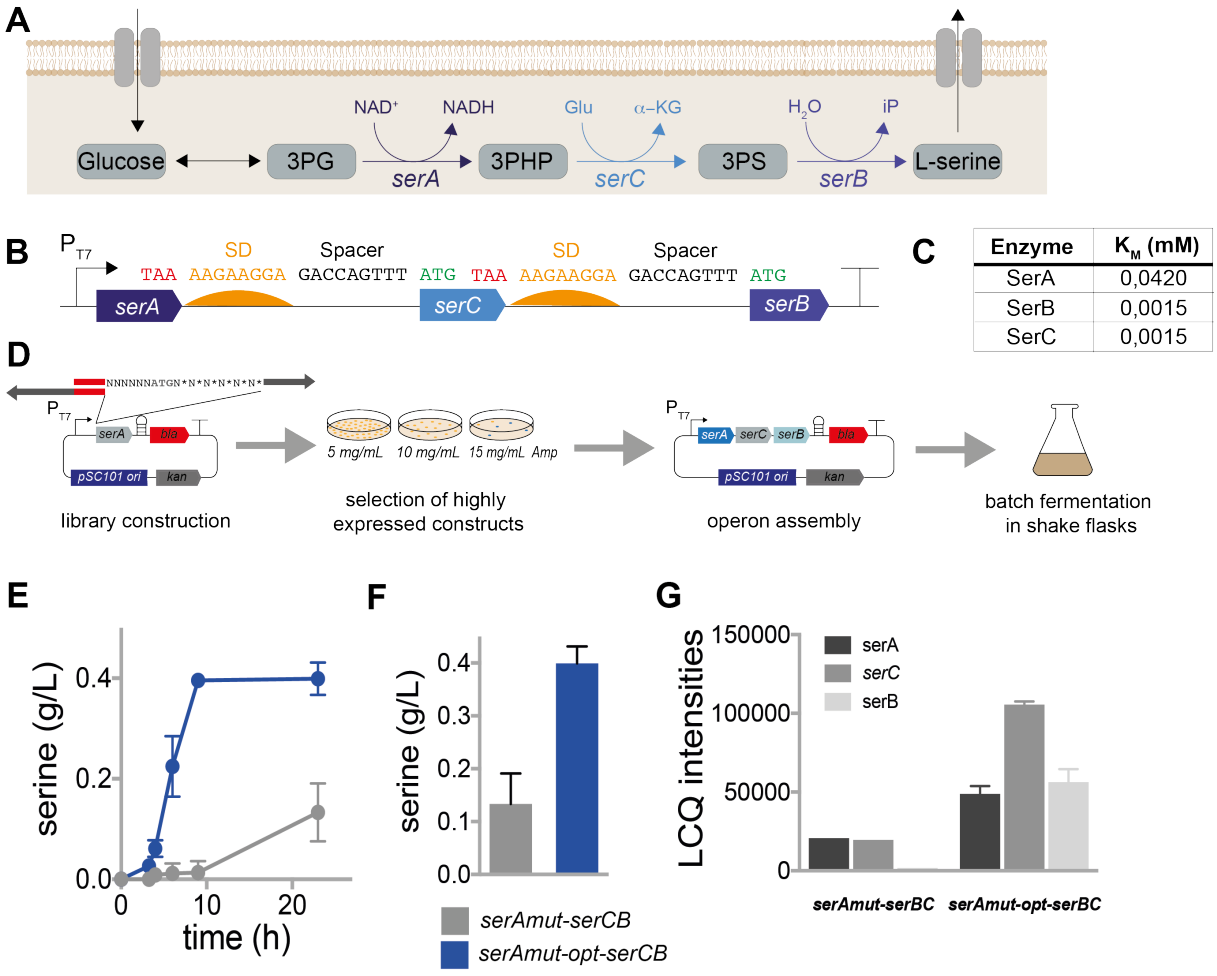
(3) Nørholm, M. H. H. (2010) A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. BMC Biotechnol. 10, 21.

(4) Hiyama, T., Nishimura, M. and Chance, B. (1969) Determination of carotenes by thin-layer chromatography. Anal. Biochem. 29, 339–42.

(5) St-Pierre, F., Cui, L., Priest, D. G., Endy, D., Dodd, I. B., and Shearwin, K. E. (2013) One-step cloning and chromosomal integration of DNA. ACS Synth. Biol. 2, 537–541.

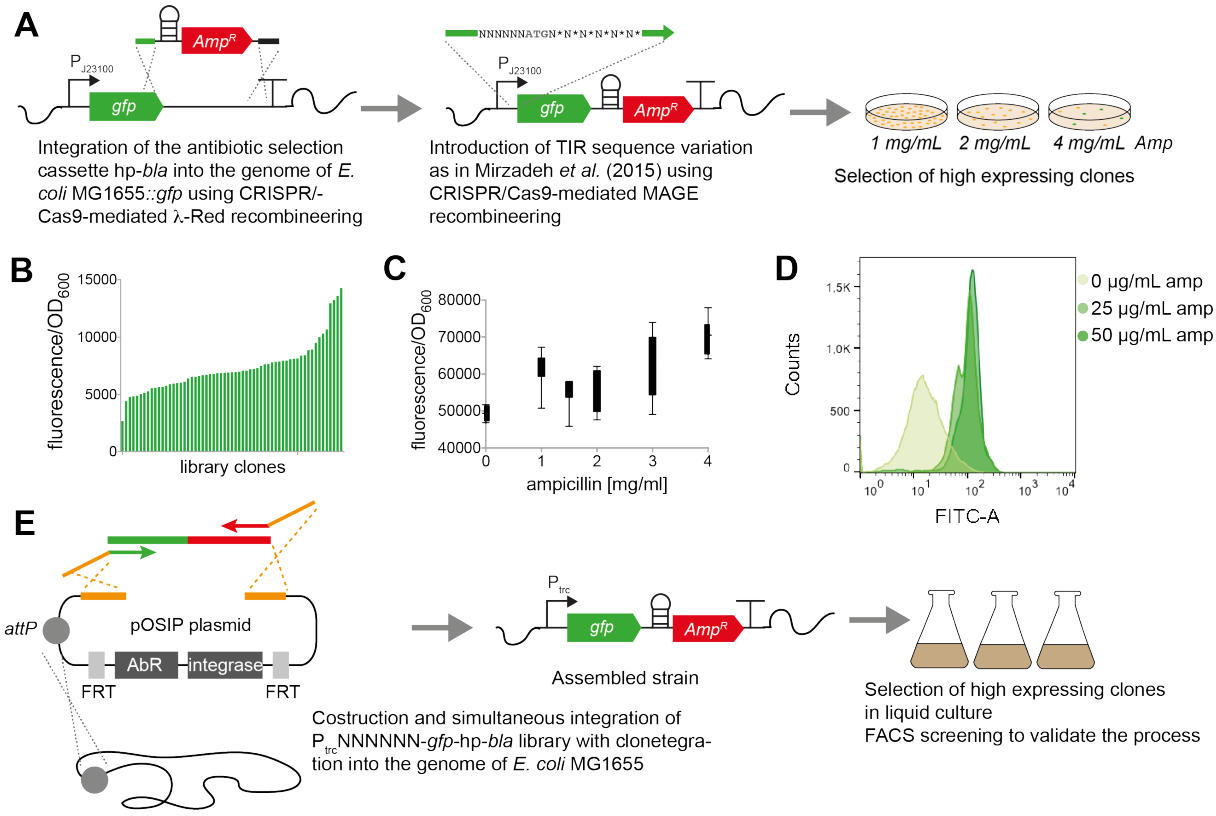
(6) Mirzadeh, K., Martínez, V., Toddo, S., Guntur S., Herrgård, M. J., Elofsson, A., Nørholm, M. H. H., and Daley, D. O. (2015) Enhanced protein production
in Escherichia coli by optimization of cloning scars at the vector–coding sequence junction. ACS Synt. Biol. 4, 959–965.

Appendix 1. Application of TIR optimization and selection for biosynthetic production pathway



Increased production of L-serine from an optimized synthetic operon in *E. coli*. (A) Glycolytic 3-phospho-glycerate (3PG) is converted in three steps into L-serine. SerA catalyzes the oxidation of 3PG to 3-phosphohydroxy pyruvate (3PHP). SerC converts 3PHP into 3-Phosphoserine (3PS). SerB catalyzes the hydrolysis of 3PS. (B) Assembly of SerACB in a synthetic operon. Start and stop codons are indicated in green and red, respectively. The Shine-Dalgarno sequence is marked in orange. (C) K_M values for SerA, SerB and SerC. Data was taken from the Sabio-RK database. SerA conducts the rate-limiting step. (D) SerAmut (feedback inhibition removed) was taken from the synthetic operon and then translationally coupled to the ampicillin resistance gene (*bla*). Sequence variation was introduced (for more details, see Chapter 3.3.4) (NNNNNNATGN*N*N*N*N*N*) and optimized variants selected on increasing concentrations of ampicillin. Hereafter, the biosynthetic pathway for L-serine was re-assembled in an operon with the optimized SerAmut as the first gene (*serAmut-opt-serCB*). (E) L-serine production was measured over time for the assembled synthetic operon (grey) and the operon with an optimized *serAmut* TIR (blue). L-serine production was measured after 3, 4, 6, 9 and 23 hours. (F) Endpoint L-serine production was compared. Optimization of SerAmut increased production titers from 0.13 g/L to 0.4 g/L, a 3-fold improvement. (G) In a proteomic analysis, expression levels of the three pathway genes in the original operon and in the optimized operon were compared. Optimized translation of the first gene, *serA* (dark grey), resulted in elevated expression of the second gene, *serC* (grey) and also in elevated expression of the third gene, *serB* (light grey). Data is derived from a manuscript in preparation, which is not included as an article in this thesis.

Appendix 2. Application of TIR optimization and selection on the genome of *E. coli*



Optimized production of GFP in genome-based expression libraries in *E. coli* (A) The gene encoding the green fluorescent protein GFP (green) was translationally coupled with the *bla* gene (red). For this reason, hp-*bla* was integrated down stream of genomically encoded *gfp* using CRISPR/Cas9 mediated λ–Red recombineering. Subsequently an expression library was constructed. Expression was controlled by a constitutive promoter (P_{J23100}). (B) A TIR library (NNNNNNATGN*N*N*N*N*N*) (for more details, see Chapter 3.3.4) was constructed with CRISPR/Cas9-mediated MAGE and 96 clones were assayed for their GFP levels to assure that sequence variation in the translation initiation region on the genome leads to differences in expression levels. (C) The P_{J23100}-GFP TIR library was grown with different ampicillin concentrations on agar plates. Seven individual colonies were picked from each plate and expression levels were estimated by fluorescence. (D) The library constructed with clonetegration was grown with different ampicillin concentrations directly in liquid culture. Expression levels of the different populations were assayed by flow cytometry. (E) A TIR library of *gfp*-hp-*bla* under control of an IPTG inducible promoter (P_{trc}) was assembled in pOSIP-KT and integrated into the genome. High expressing variants were then selected in liquid cultures with increasing ampicillin concentrations. Data is derived from the patent application EP17186000.0.